

**REDISTRIBUTION OF STATES AND INDUCING
NEW CHANNELS FOR CONFORMATIONAL CHANGE:
COMPUTATIONAL STUDIES ON CALMODULIN**

by
AYŞE ÖZLEM AYKUT

Submitted to the Graduate School of Engineering and Natural Sciences
in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

Sabanci University
Spring 2013

© Ayşe Özlem Aykut 2013
All Rights Reserved

REDISTRIBUTION OF STATES AND INDUCING
NEW CHANNELS FOR CONFORMATIONAL CHANGE: COMPUTATIONAL
STUDIES ON CALMODULIN

APPROVED BY

Prof. Canan Atılgan
(Thesis Supervisor)

.....
Canan Atılgan

Prof. Ali Rana Atılgan
(Thesis Co-Supervisor)

.....
Ali Rana Atılgan

Prof. Pemra Doruker Turgut

.....
Doruker Turgut

Assist. Prof. Elif Özkırımlı Ölmez

.....
Elif Özkırımlı Ölmez

Assist. Prof. Deniz Sezer

.....
Deniz Sezer

DATE OF APPROVAL: 10.06.2013

REDISTRIBUTION OF STATES AND INDUCING
NEW CHANNELS FOR CONFORMATIONAL CHANGE: COMPUTATIONAL
STUDIES ON CALMODULIN

AYŞE ÖZLEM AYKUT

Materials Science & Engineering, Doctor of Philosophy Thesis, 2013

Thesis Advisors: Prof. Canan Atilgan, Prof. Ali Rana Atilgan

Keywords : Molecular-Dynamics Simulations; Calcium-Loaded Calmodulin; Central Helix; Proteins; Spectroscopy; Algorithm; pH; Plasticity; Elastic Network Models; Interdomain Interactions; Macromolecular Motions; Multidimensional NMR

Abstract

In vitro experiments demonstrate that large conformational changes in many proteins are observed as “rare events” occurring on microsecond timescales. For proteins that sustain a plethora of functions, it is imperative that different conformational states be achieved readily under slightly differing environmental conditions *in vivo*. We investigate how perturbations that may be experienced by proteins in their fluctuating environments may be invoked to facilitate their access to different micro states, using the example of calcium loaded calmodulin (Ca^{2+} -CaM). As perturbations, we introduce (i) mutation of a single residue, (ii) protonation of a group of residues to mimic the low pH environment, (iii) external forces applied on single residues, (iv) external displacements applied on single residues, and (v) increasing the ionic strength (IS) of the solution the protein resides in. After performing molecular dynamics simulations on the perturbed systems (i) and (ii), we observe distinct conformational changes within tens of nanoseconds, that otherwise occur on the time scales of microseconds. In particular, a reversible change between the extended and compact Ca^{2+} -CaM structure may be invoked via the E31A mutation. This compact form bears a bent linker which is observed in many of the ligand bound forms of Ca^{2+} -CaM. Protonation of ten acidic residues also leads to a large conformational change on the time scale of 100 ns. The structure attained is compact and although it does not have

a bent linker, it is compatible with fluorescence resonance energy transfer and nuclear magnetic resonance experimental data. Barrier crossing between extended and compact forms of CaM, which is normally a rare event, is facilitated by shielding the repulsive electrostatic interactions between the two lobes. This is due to either the impact of lowering the pH in the environment or to allosteric interactions originating on the N-terminal domain and detected in the C-terminal domain as implicated by from the results of (iii) and (iv). We further find using scenario (v), at high IS there is depletion in the number of ions residing in close proximity of the protein. The loss of ionic screening results in rigidification of the linker fluctuations, leading to the slowing down of the dynamics of the system. Application of external perturbations to extended Ca^{2+} -CaM in the form of forces (iv) as opposed to displacements (v) yield complementary results. We find that both approaches designate the same two regions on the protein structure, making these regions potential sites for manipulating conformational change. Local force perturbations implicate charged residues while local displacement perturbations find polar and hydrophobic residues in the same two regions. The observations reflect the differences inherent in the thermodynamic functions optimized by the two approaches, while confirming the universality of allosteric communication between the two lobes.

KONFORMASYONAL YAPILARIN DAĞILIMININ DEĞİŞTİRİLMESİ VE
YAPISAL DEĞİŞİM İÇİN YENİ KANALLARIN UYARILMASI:
KALMODULİN ÜZERİNDE HESAPLAMALI ÇALIŞMALAR

AYŞE ÖZLEM AYKUT

Malzeme Bilimi ve Mühendisliği, Doktora Tezi, 2013

Tez Danışmaları: Prof. Dr. Canan Atılğan, Prof. Dr. Ali Rana Atılğan

Anahtar Kelimeler : Protein nanomekaniği, biyofizik, moleküler mekanik, enerji yüzeyi, protein-protein etkileşimleri, elastik ağyapı modelleri, moleküler dinamik, kalsiyum yüklü kalmodulin, makromoleküler hareketler, çok boyutlu NMR

Özet

In vitro deneyleri proteinlerin büyük yapısal değişim geçirebilmelerinin “nadir bir olay” olduğunu ve bu olayların zaman ölçeğinin microsaneye olduğunu göstermiştir. Birden çok fonksiyonu gerçekleştirebilen proteinlerin, çevresel koşullarının değiştirilmesi ile *in vivo* ortamda yapısal değişimi kolaylıkla geçiriyor olabilmesi zorunludur. Bu çalışmada kalsiyum yüklü kalmodulin (Ca^{2+} -CaM) örneğini kullanarak, proteinin etkileşim ortamında uygulanan zorlamalara verdiği tepki ile başka bir konformasyona geçişinin kolaylaştırılmasını inceliyoruz. Zorlama olarak, (i) tek bir amino asidin mutasyonu (ii) düşük pH etkisini göstermek için bir grup amino asidin protonlaması (iii) sistematik olarak seçilen amino asitlere sarsım verilmesi (iv) sistematik olarak seçilen amino asitlere öteleme verilmesi (v) proteinin içinde bulunduğu solventin iyonik kuvvetini arttırmak olmak üzere beş farklı tip kullanılmıştır. Zorlama uygulanmış sistemlerden (i) ve (ii)’nin moleküler dinamik benzetimleri yapılmış ve deneysel koşullarda microsaneye zaman ölçeğinde gözlemlenen değişimin onlarca nanosaniye zaman ölçeğinde gerçekleştiği gözlemlenmiştir. Özellikle, E31A mutasyonu ile açık ve kapalı yapılar arasında tersinir geçiş sağlanabilmiştir. Bu yapı, bağlantı helisinin bükülmesi ile kapalı hali almıştır ve bu durum ligand bağlı yapılarda gözlemlenen ile aynıdır. Bir grup amino asidi protonlamak da 100 ns zaman ölçeğinde gerçekleşebilen bir yapısal değişime yol açmaktadır. Bu durumda ulaşılan yapı kapalıdır ve bağlantı helisinde

bükölme gözlenmemiştir. Ayrıca, bu yapı floresans rezonans enerji transferi (FRET) ve nükleer manyetik rezonans (NMR) deneysel verilerinden elde edilmiş yapılar ile uyumludur. Açık ve kapalı yapılar arasındaki enerji bariyerinin geçilebilmesi ki deneysel koşullarda “nadir bir olay” olarak tanımlanır, iki lob arasındaki itici elektrostatik etkileşimler kapatıldığı için gerçekleşebilmiştir. Bu değişim, ya düşük pH daki etkinin bir sonucu ya da N lobda bulunan amino asitler arası oluşan alosterik etkileşim ile C lobu etkilemesi sonucu olabilmektedir ((iii) ve (iv) zorlalarının sonucu ile de gösterilmiştir). En sonuncu zorlama senaryosu olan (v), protein’in bulunduğu ortamın iyonik kuvvetinin arttırılması durumunda, protein’e yakın olarak duran iyon sayısında azalma olduğu gözlemlenmiştir. Bu durum bağlantı helisinde bulunan amino asitlerin salınımlarını sertleştirmiş ve protein yapısının salınım dinamiğini yavaşlatmıştır. Ca^{2+} -CaM açık yapısında sistematik olarak amino asitlere sarsım uygulanması (iii) kuvvet olarak ya da (iv) öteleme olarak birbirini tamamlayıcı sonuçlar vermiştir. İki yaklaşımda proteinin yapısal değişimini tetikleyen amino asitlerin protein üzerinde aynı bölgelerde olduğunu göstermektedir. Yerel kuvvet metodu yüklü amino asitleri, yerel öteleme metodu ise polar ve hidrofobik amino asitleri aynı bölgede bulmaktadır. Her iki yaklaşımda yapılan gözlemler iki lob arasında alosterik iletişim evrenselliğini teyit ederken, iki metodun farklı tipteki amino asitleri bulması metodların optimize termodinamik fonksiyonlarındaki farklılıkları yansıtmaktadır.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank to my supervisors Canan Atılgan and Ali Rana Atılgan. This thesis may have not been possible without their support. Not only our valuable academic discussions, also the encouragements they provided me through my academic and social life help me to find the right way always easier. I would also like to Deniz Sezer, for the fruitful discussions and significant remarks. I would like to thank Osman Uğur Sezerman for being the first academic person in my life, as my uncle, and driving me towards academia and for his endless support and motivation. I would like to thank Torsten Schwede for giving me the opportunity to join his group as a visiting student and providing me a scientific environment in Basel.

We had successfully created a great office atmosphere in our FENS 2107 lab that brings enjoyable working environment for which I like to present here my appreciation. Special thanks go to Irmak Sirer, Deniz Turgut, Ibrahim Inanç, Murat Mülâyim, Gökhan Kaçar, Gökçe Güven, Kaan Bilge, Elif Özden Yenigün and Haleh Abdizadeh. Also, I would like to thank all my friends from MAT and BIO program: Aslı Yenigün, Burcu Saner Okan, Lale Tunçyürek, Firuze Okyay, Özlem Kocabaş. I would like to thank to all of the group members in Torsten Schwede Lab. Especially to Valentina Romano, Tobias Schmidt, Marco Biasini and Juergen Haas for their helpful discussions, friendship and providing a cozy working environment.

At last but not the least, I would like to thank to my family for their great guidance, supports and patience during this study. To my husband and my daughter, without whom, this work could not be made possible.

Table of Contents

List of Figures.....	xi
List of Tables.....	xiv
List of Abbreviations.....	xv
CHAPTER 1. Introduction.....	1
1.1. Overview.....	1
1.2. Background.....	3
1.2.1. Probes of CaM Structure and Function.....	3
1.2.2. Techniques Probing Computational Multiplicity.....	6
1.3. Motivation.....	9
CHAPTER 2. Theoretical and Computational Methods	10
2.1. Protein Study Set	10
2.2. System Construction and Molecular Dynamics Simulation Protocol	12
2.3. Local Perturbations Introduced by Perturbation Response Scanning (PRS).....	14
2.3.1. Linear Response Theory.....	14
2.3.2. Perturbation Response Scanning (PRS) Protocol.....	18
2.4. Local Perturbations Introduced by Molecular Mechanics (MM).....	18
2.5. Steered Molecular Dynamics (SMD) Simulations.....	21
2.6. Quantifying Structural Similarity by Overlap.....	22
2.7. Reduced Degrees of Freedom of CaM	23
2.8. Inter - Helical Angles	24
2.9. pKa Calculations.....	24
CHAPTER 3. Results.....	26
3.1. Structure and Sequence Analysis of the Proteins.....	26
3.2. Perturbation Response Scanning Analysis.....	33
3.3. Molecular Dynamics Simulations.....	33
3.3.1. MD Simulation Trajectories of the 3CLN and 1PRW Structures (3CLN, 1PRW).....	33
3.3.2. MD Simulation Trajectories of 3CLN with 10 Residues Protonated (3CLN*).....	37

3.3.3. MD Simulation Trajectories of 3CLN System; only E31 Protonated (E31 ⁺).....	40
3.3.4. MD Simulation Trajectories of 3CLN System; E31 mutated (E31A)	42
3.3.5. MD Simulation Trajectories of 3CLN System in 150 mM IS.....	49
3.4. Response of CaM to Local Displacements.....	51
3.5. Constant Force Perturbation Introduced by Steered MD.....	54
CHAPTER 4. Discussion.....	58
4.1. Features of the Conformational Change Revealed by PRS.....	58
4.2. E31 is a Signaling Residue for Global Communication in CaM.....	59
4.3. Role of Electrostatics on the Conformational Change of 3CLN*.....	64
4.4. Comparison of the Local Perturbations given by PRS and MM.....	68
CHAPTER 5. Conclusion and Future Work.....	70
5.1. Major Findings and Contribution.....	70
5.2. Future Work.....	72
Appendix A. Time Evolution of End-to-End Distance from the MD Simulations.....	73
Appendix B. Time Evolution of Torsion Angle from the MD Simulations.....	76
Appendix C. Molecular Mechanics Script.....	79
References.....	81
List of Publications.....	89

LIST OF FIGURES

Figure 1. Three dimensional representations of the extended (3cln) and compact (1prw) forms of Ca^{2+} -CaM. (NTD is green, linker is magenta and CTD is cyan, gray spheres are Ca^{2+} ions)

Figure 2. Three-dimensional structures of the proteins studied in this work. The initial structure is placed in the center, and the targets are oriented in such a way that their C-terminal domains are best fitted. N-terminal domain is in green, C-terminal domain is in cyan, and the linker is in magenta. The Ca^{2+} ions are shown as gray spheres, the bound ligand molecules are shown in gray surface representations.

Figure 3. Free-body diagram representation of a residue.

Figure 4. Pulling in one dimensional case. Dummy atom and SMD atom are colored in red and blue, respectively.

Figure 5. l_{linker} vs φ plot representing CaM structures in the study set. Black full circles represent the 2k0e NMR ensemble. Open circles are the X-ray structures in the study set.

Figure 6. Degree of ionization as a function of pH for (a)Asp/Glu (36 residues) and (b)Lys/Arg (13 residues) amino acid types. E31 (dashed) and D122 (gray) are distinguished as those capable of changing ionization with subtle pH variations at physiological conditions. The standard pKa values for non-perturbed residues are 4.4 for Asp/Glu, 10.0 for Lys, 12.0 for Arg [1].

Figure 7. RMSD of the protein trajectory for 3CLN; moving average taken over 200 data points of the RMSD of the whole protein shown in black, NTD shown in green, CTD shown in cyan and linker shown in purple. The light gray lines show the actual RMSD of the full trajectory.

Figure 8. Representative snapshots taken from different parts of the 3CLN MD trajectory. The number representing each part is given at the top.

Figure 9. φ vs l_{linker} joint probability plot for 3CLN trajectory.

Figure 10. RMSD of the protein trajectory for 1prw (two independent MD trajectories, each 10000 snapshots); moving taken over 100 data points of the RMSD of the whole protein shown black, NTD shown green, CTD shown cyan and linker shown purple. The light gray lines show the actual RMSD of the trajectory.

Figure 11. l_{linker} vs φ joint probability plot for 1prw runs. (a) 1PRW_r1 (b) 1PRW_r2.

Figure 12. RMSD of the protein trajectory for 3CLN*; moving average taken over 100 data points. RMSD of the whole protein shown black, NTD shown green, CTD shown cyan and linker shown purple. The light gray lines show the actual RMSD of the trajectory.

Figure 13. l_{linker} vs φ joint probability plot for 3CLN* run.

Figure 14. Schematic representation of the energy landscapes discussed according to the torsional angle change. The dotted lines show the energy level of the relaxed systems at the third part of the corresponding simulations.

Figure 15. RMSD of the protein trajectory for E31⁺ (two trajectories of 15000 and 20000 snapshots, respectively) ; moving average of the RMSD of the whole protein shown black, NTD shown green, CTD shown cyan and linker shown purple. The light gray lines show the actual RMSD of the trajectory. Moving average taken over 150 and 200 data points, respectively.

Figure 16. l_{linker} vs φ joint probability plot of E31⁺ run.

Figure 17. RMSD of the protein trajectory for E31A (three trajectories of each 20000 snapshots); moving average taken over 200 data points for each trajectory. RMSD of the whole protein shown black, NTD shown green, CTD shown cyan and linker shown purple. The light gray lines show the actual RMSD of the trajectory.

Figure 18. l_{linker} vs φ joint probability plots of E31A runs.

Figure 19. Representative snapshots taken from different parts of the E31A MD trajectory. The number representing each part is given at the top.

Figure 20. The dark gray points are the median of the pathways generated by FRODA and the light gray shaded area is the standard error bars along each axis. The initial (50 ns) and target (80 ns) structures that are fed to Geometrical Pathways are shown with black outlined empty circles. . Forward transition (64.3 – 66.7 ns) and reverse transition (145.2 – 147.2 ns) paths.

Figure 21. Inter helix angle trajectories. In the upper panel, angle trajectories for the NTD in each run are shown. The bottom panel shows angle trajectories in CTD.

Figure 22. The angles between the helices in consecutive EF Hands throughout each MD run. Top panel shows the ones for NTD and bottom panel shows the ones for CTD.

Figure 23. RMSD of the protein trajectory for 3CLN^{150mM} (trajectory of 7500 snapshots) ; moving average taken over 75 data points. RMSD of the whole protein shown black, NTD shown green, CTD shown cyan and linker shown purple. The light gray lines show the actual RMSD along the trajectory.

Figure 24. l_{linker} vs φ joint probability plot of 3CLN^{150mM} run.

Figure 25. Angles between the helices in the EF Hands and between the consecutive EF hands throughout the MD trajectory of 3CLN at 150 mM IS. Top panel shows the ones for NTD and bottom panel shows the ones for CTD.

Figure 26. Molecular Mechanics perturbation response maps of 3CLN for different directions (a, b, c) are shown. The positions of the secondary structural units are shown along the sides; red indicates α -helix, yellow arrow indicates β -sheet, and purple indicates loops.

Figure 27. (a) Best PRS prediction of the displacement vectors (green) belonging to the 3cln to 1lin conformational change, overlaid on the initial structure. The main motion is a bending of the two lobes accompanied by rotation of the linker, where motions are especially accentuated in the region of residues 75–90. (b) Coordination of the E31 related Ca^{2+} ion. The side-chain (χ_1, χ_2) angles of E31 are in (t, t) conformation. Two alternate conformations of E31 which do not clash with any other heavy atoms in this conformation are also shown as transparent traces. These have (t, g^-) and (g^-, g^+) conformations for the (χ_1, χ_2) angle pair. The thick red arrow represents the best perturbation direction of E31 in both figures. (c)

Figure 28. Snapshots of the compact structures accomplished by E31A_r1 (80ns, cyan) and SMD (2ns, orange).

Figure 29 . Comparison of the structure attained in 3CLN* run 2 with the best-fitting NMR compatible one from the PDB file 2K0E. with 4.6 Å RMSD of the overall backbone, 2.5 Å in the N-lobe and 3.7 Å in the C-lobe.

Figure 30. 3D representation of 3cln where the residues identified shown in surface representation for MM in silver and for PRS in orange on the protein.

Figure 31 . Nonbonded interaction energy between NTD and CTD in the 3CLN* MD simulation trajectory. Most contributions come from the electrostatic component (dashed line), displaying that the van der Waals term has negligible effect on the conformational change.

Figure 32. The radial distribution between the heavy atoms on 3cln and the ions Na^+ ions (as well as Cl^- ions for the 150 mM IS case) in the solvent for 76 mM IS and 150 mM IS conditions.

LIST OF TABLES

Table 1. Calmodulin structures studied and their properties.

Table 2. Features of the protein-water systems used in the MD simulations.

Table 3. RMSD between pairs of structures listed in Table I. Lower diagonal: RMSD between overall structures; upper diagonal: RMSD between NTD (**bold**) and CTD (*italic*) only. Overlap between the experimental displacement vectors $\Delta\mathbf{S}$ are also displayed in parentheses.

Table 4. Best overlap values obtained for proteins studied by PRS.

Table 5. pKa values calculated for the structure 3cln by various methodologies; results from other computational (MCCE) and, where possible, experimental work is also included.

Table 6. Average inter-helical angles of helices V-VI and standard deviations.

Table 7. Best overlap values obtained for proteins studied by MM .

LIST OF ABBREVIATIONS

CaM	Calmodulin
Ca ²⁺ -CaM	Calcium loaded Calmodulin
IS	Ionic Strength
MD	Molecular Dynamics
SMD	Steered Molecular Dynamics
PRS	Perturbation Response Scanning
MM	Molecular Mechanics
NTD	N Terminal Domain
CTD	C Terminal Domain
RMSD	Root Mean Square Deviation

CHAPTER 1

Introduction

1.1. Overview

The living cell is a complex system comprising thousands of biological macromolecules which constantly interact with each other to maintain the function of the cell, reflecting the dynamics of cellular networks [2, 3]. A subset of these macromolecules is comprised of proteins which are the key actors of many biological processes. Most of these processes involve interactions of proteins with each other and these interactions are required to be performed on fast time scales to allow the cell to respond to different stimuli [2, 3]. A protein may have several interaction partners that vary with the type of stimulus and the biological process to be carried out [4]. In most cases the protein needs to adapt to a specific configuration to be able to interact with its partner for that process [5, 6].

Hence, most of the proteins in the cell have diverse functionality related to their ability to change between distinct conformations available along their landscape [7-10]. The mechanisms of conformational transitions that accompany the process of complex formation, ionization changes due to the change pH and ion concentration, as well as the temperature change by proteins have been the subject of intense experimental and theoretical research for several decades [6, 11-16].

Understanding the mechanism of the proteins' ability to span diverse configurations would allow regulating and modifying these processes which may further be used for design purposes. One may designate mutations that may change the functionality of proteins or design drugs that may interfere with the key sites crucial for the conformational changes achieved by that protein.

In this study, we aim to provide a basis for the significantly large conformational changes in proteins occurring in response to perturbations arising from environmental factors, such as change in pH or ionic strength, or that arising from local factors, such as

mutation of a residue or change in the protonation state of a residue [17, 18]. We focus on a Ca^{2+} sensor protein, calmodulin (Ca^{2+} -CaM) which plays a crucial role in various cellular signaling cascades through regulation of numerous target proteins in a Ca^{2+} dependent manner while going through diverse configurations to interact with its diverse partners [19-21]. In light of the critical regulatory role of CaM in binding many proteins to trigger their functions, it is crucial to unravel the key residues that may be targeted to facilitate the conformational transition. It is known that comparison of experimentally determined ligand bound/unbound forms of a protein gives a wealth of information on the basic motions involved, as well as the residues participating in the functionality [22-25]. In this study we utilize the unliganded and liganded conformations of Ca^{2+} -CaM [26-33]. Due to the time scales involved ($100\ \mu\text{s}$; see ref [34]) that are much slower than may be observed by standard molecular dynamics (MD) simulations, other computational and theoretical approaches must be employed. We therefore use in-house developed coarse grained methods [22, 35] in conjunction with extensive MD simulations. We aim to understand the extent of applicability of linear response theory [36] to unliganded form of CaM, and we study conformational motions utilized to achieve the ligand bound forms of different CaM structures, where the size of the ligands varies from a few atoms to peptides of 26 residues long [17]. We also investigate the range of conformations sampled by Ca^{2+} -CaM using extensive MD simulations. We make a systematic study of the dynamics of Ca^{2+} -CaM on time scales which are at least 150 ns for seven separate systems that have one of the following: (i) a local perturbation in the form of a mutation in a single residue; (ii) a local perturbation in the form of a protonation in a residue; (iii) a change in the ionic strength (IS); (iv) a change in pH; and (v) starting with an initial conformational state alternative to that observed in most crystal structures.

The methods developed in this thesis may be applied to all proteins that are known to span several configurations to perform their functions, such as maltose binding protein, adenylate kinase (ADK), immunoglobulin and GroEL [37]. One may also study the impact of mutations, environmental factors such as pH and IS to the protein equilibrium dynamics and to the shifts in its conformational space [38-43].

The thesis is organized as follows: Chapter 1, **Introduction** describes the problem we aim to tackle and its biological relevance. The second section in Chapter I provides the background, giving a brief description of CaM followed by a review of computational studies conducted to understand the dynamical behavior of this protein in

relation to its functionality. In Chapter 2, **Theoretical and Computational Methods**, we describe the structures studied, the perturbation response scanning (PRS) technique and the theory behind, the molecular mechanics (MM) methodology, the MD simulations and steered MD simulation protocols performed. The pKa calculation methods used in this study are also summarized therein. In Chapter 3, **Results**, the structure and sequence analysis of the CaM structures are given to discern the types of conformational changes involved. Also describes in this chapter are the results obtained by application of coarse grained methods and the MD simulations described in Chapter 2. In Chapter 4, **Discussion**, first features of the findings revealed by PRS are discussed. The role of electrostatic interactions for the multiple residue protonated CaM is discussed in detail. Comparison of PRS and MM techniques are shown. In Chapter 5, **Conclusion and Future Work**, the important findings are outlined in the frame of single residue perturbations and electrostatic environment changes. Also describes the future work that can be conducted with utilization of the results from this study.

1.2. Background

One of the notorious examples of proteins that can take on many different conformations in their functional state is CaM. It has a pivotal role as an intracellular Ca^{2+} receptor that is involved in calcium signaling pathways in eukaryotic cells. It has been shown that CaM binds and regulates more than 300 target proteins (ligands) and that its structural plasticity is crucial for enabling its interaction with its diverse partners [20]. Binding of Ca^{2+} and proteins or small organic compounds to CaM induces large conformational changes that are distinct to each interacting partner [20, 21, 32]. The interaction of CaM with target proteins at various levels of Ca^{2+} loading control many key cell processes as diverse as gene expression, neurotransmission, and ion transport [44]. Also, diseases related to unregulated cell growth, such as cancer, have been shown to have elevated levels of Ca^{2+} -CaM [45]. Structural heterogeneity of CaM depends significantly on the nature of the ligand, environmental conditions of pH, ionic strength and temperature [34, 46-48].

1.2.1. Probes of CaM Structure and Function

CaM consists of two homologous domains, the N-terminal domain (NTD) and the C-terminal domain (CTD), which are separated by an interdomain linker. Each domain is composed of two EF-hand helix-loop-helix motifs, with a highly conserved

amino acid sequence in eukaryotes. While performing its function, CaM undergoes significant conformational changes concerning both the EF-hand motifs themselves and the mutual orientation of the lobes. These motifs occupy a “closed” conformation in the calcium-free state, in which the helices in the two pairs of EF hands are closely packed together, Ca^{2+} ligation leads to an “open” state in which significant changes in conformation in each EF-hand pair result in the exposure of a hydrophobic cleft in both domains (Figure 1) [49, 50].

The exposure of the hydrophobic cleft increases the affinity of CaM to a wide range of binding partners [31, 51, 52]. In addition to these, at low Ca^{2+} levels, it was found that the two domains act uncoupled, while stabilization of especially the CTD upon Ca^{2+} loading leads to coupled motions between the domains, with concerted rotational dynamics occurring on the order of 15 ns time scale [53].

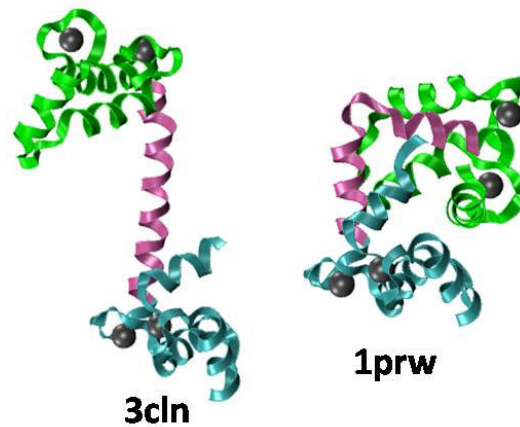


Figure 1. Three dimensional representations of the extended (3cln) and compact (1prw) forms of Ca^{2+} -CaM. (NTD is green, linker is magenta and CTD is cyan, gray spheres are Ca^{2+} ions).

The linker between the NTD and CTD plays a key role in the function of CaM. The coupling between the lobes is orchestrated by this flexible linker region, letting the two domains adopt a large distribution of relative orientations so that its wide variety of different targets may be accommodated [54, 55]. Although this linker is seen as a long exposed helix in the crystal structure of Ca^{2+} -CaM [56], solution-state nuclear magnetic resonance (NMR) relaxation measurements [54] reveals that its central region is highly flexible in the Ca^{2+} -CaM state. Moreover, NMR experiments also point out that Ca^{2+} -CaM adopts a distribution of conformations [57], whereby neither the originally observed dumbbell shaped [26] nor the later recorded much compact crystal structures [27], also presented in Figure 1, are in abundance in solution. Single molecule

experiments have also established a distribution of possible structures and how they shift with changes in environmental conditions such as Ca^{2+} concentration, pH and/or ionic strength [46]. Dynamical information obtained from fluorescence resonance energy transfer (FRET) experiments measuring the distance distributions between labeled sites residing on separate lobes of Ca^{2+} -CaM show that not only do at least two conformations exist in solution under physiological conditions, but that the equilibrium is shifted toward a single, more compact form upon decreasing pH from 7.4 to 5.0 [47, 57]. More recent fluorescence correlation spectroscopy experiments have quantified the time scale of interconversions between the various available states to be on the order of 100 μs [34]. Other experimental work also investigates the dynamic nature of CaM conformations; e.g. 1.0 Å crystal structure of Ca^{2+} -CaM provides evidence for domain displacements and discrete backbone disorder, suggesting that it occupies a number of hierarchical set of substates even in the crystal environment [56].

The importance of the linker flexibility becomes apparent on examining the structures of Ca^{2+} -CaM in complex with target peptides [58]. It was shown that negatively charged side chains in Ca^{2+} -CaM are being attracted to positively charged residues in many of its targets [59]. In most structures of known complexes, the NTD and CTD are effectively clamped together around a helical peptide in the compact conformation, although other studies reveal that a variety of other binding modes exist. One example of the latter is the Ca^{2+} -CaM extended binding conformation, in which the linker stays in the long helix form resulting NTD and CTD to be apart as is the case in small conductance potassium channel [60] or adenylate-cyclase [61]. Even after target recognition, CaM does not adapt to the well-known compact binding conformation in these cases.

All these experimental data identify CaM as a protein which samples multiple conformations and that can shift between these conformations. It adopts a variety of conformations upon binding to target molecules, or as a response to changing environmental conditions even in the absence of binding to a target molecule. Many of these conformations are revealed from solution techniques such as small angle X-ray scattering (SAXS) and nuclear magnetic resonance (NMR) spectroscopy. These techniques provide experimental observables that are averages over a manifold of conformations with different weights. Conversely, the conformations revealed by X-Ray crystallography usually represent a single conformation trapped in the crystal. Bertini and coworkers introduced the concept of maximum occurrence (MO) i.e., the

maximum percent of time that flexible proteins can spend in any given conformation [62]. They applied this idea to CaM using SAXS and NMR data (using psuedocontact shifts and self-orientation residual dipolar couplings arising from the presence of paramagnetic lanthanide ions) and found that “closed” and “fully extended” conformations trapped in the crystalline forms of CaM have MOs of only 5 and 15 %, respectively [62]. Other compact conformations in general have small MOs, whereas some extended conformations have MO as high as 35%, strongly suggesting these conformations to be most abundant in solution [62]. A dynamic population shift model has been proposed to explain the structural transitions observed during complex formation in terms of a variation in the equilibrium distributions of pre-existing populations that interchange dynamically in the absence of a binding partner [5, 63-67]. This model suggests that the protein continuously samples a range of substates whose statistical weights are redistributed (i) upon a perturbation in the environment – reflected by the change in pKa’s of the residues and, (ii) a perturbation in the structure – reflected as a mutation or protonation of residue/s. The MO concept supports this model as multi-conformational CaM samples substates that have different MOs in the Ca^{2+} -CaM state.

1.2.2. Techniques Probing Computational Multiplicity

A variety of computational techniques have been developed to distill information from the Protein Data Bank (PDB) structures [68], going beyond the original studies that relied on their visual inspection. Normal mode analysis, in particular, turned out to be a useful technique in the analysis of functionally related conformational dynamics [69, 70]. As such, it has made possible the classification of protein motions, e.g., hinge bending and shear [71]. It has further been shown that the predominant contributions to these motions may be described by a single, most collective mode for some proteins, whereas it may be obtained from a superposition of several modes for others [72]. With the advent of coarse graining of biomolecular structures through residue-based network models [71, 73-76], it has been possible to study a large number of protein structures. Statistical mechanical connection relates residue-based network properties to experimentally determined fluctuation patterns of residues [73], in turn allowing separation of contributions from fast and slow modes of motion to equilibrium dynamics [77]. These anisotropic network models (ANMs) take into account the three-dimensional geometry of interacting pairs of residues to study the

modal behavior of proteins. Using such information, it is possible to morph between the unbound and bound structures to gain insight into the intermediates that lead to the final state [25, 37, 78-82]. These analyses also uncover the different modes stimulated by different ligands binding to the same unliganded form. Despite its success in identifying the types of motions involved in ligand binding, ANM is an equilibrium methodology, providing information on the equilibrium fluctuations and the various contributions to those fluctuations from different modes of motion.

Although there is considerable success in predictions made by these analyses, there is also a debate on and lack of insight into how these modes are used by different types of proteins or by different ligands acting on the same protein [25, 37, 78-82]. It is also unclear how many modes are stimulated upon binding of ligands. In particular, if the conformational change involved is more complicated than, e.g. hinge bending type motion of domains, several modes may be operational at the same time to recover the motion observed. Collectivity of the modes is detrimental to the ability of representing the motion by a few slow modes [75, 83] .

It is utmost interest to use a methodology that puts the system slightly out of equilibrium and monitors the evolution of the response so as to get useful information similar to those obtained in experiments. Experimentally, the perturbation given to the system may be in the form of changing the environmental factors [84], or may target specific locations on the structure itself [85, 86]. Theoretically, the perturbation might be a force given to the system mimicking the above-mentioned forms. The response of the system is then recorded to detect the underlying features contributing to the observations, thus yielding information additional to that of the operating modes of motion.

It is possible to find examples of studies comprising of such approaches in the literature, all operating in the linear response regime. In one all-atom study, the perturbation has been applied as a frozen displacement to selected atoms of the protein, followed by energy minimization; the response has been measured as the accompanying displacements of all other atoms of the protein [35, 87]. This method, based on MM, scans all of the residues to produce comparative results. In various studies, the perturbations on residues are introduced by modifying the effective force constants [88], links between contacting residue pairs [82, 89] or both [4]. On the other hand, perturbations may be inserted on the nodes instead of the links between pairs of nodes; depending on the location of the perturbation, the resulting displacements between the

unbound and bound forms may be highly correlated with those determined experimentally [36].

The PRS method has been developed [22] with the assumption that the equilibrium fluctuations at a given local energy minimum of the protein has the tendency to convert to other viable conformations via paths that may be approximated linearly. Thus, characterizing the fluctuations by the Hessian matrix of the protein, single residue perturbations are inserted as forces and the resulting conformational change is calculated using linear response theory [36]. The protein is scanned residue-by-residue by insertion of force to each in hundreds of possible random directions. The residue/force combinations are determined that best represent the conformational change between open/closed forms of the protein under investigation. We found that in many cases, the response to single residue perturbation represents the conformational change and may be used to uncover allosteric pathways [23].

The ferric binding protein (FBP) was studied in detail by using PRS and it has been possible to map those residues that are structurally amenable to inducing necessary conformational change [22]. The main assumption is that for some conformational transitions, linear response assumption may be valid so that it is possible to perturb a structure residing in one of the basins of its free energy landscape and find the route that connects that basin to another one using linear response theory.

As for CaM dynamics, it has been studied by using MD with different aspects. The fully loaded CaM structure was investigated under various force fields, different water models and at different ionic strengths [90-92]. In all these studies, the flexibility of the linker was reported as the most important result, postulated to lead to the reorientation of the two domains into a position favorable for target binding. ^{15}N NMR measurements of the backbone dynamics of Ca^{2+} -CaM indicate a very high degree of mobility for the linker residues 78-81 [54]. As a result of the conformational variability of the linker, a decrease in the radius of gyration was observed at 60 mM IS [93]. Another study showed that, Ca^{2+} -CaM changes conformation when introduced to a solvent at low pH and low ionic strength (IS). Also, it was proposed that electrostatic interactions between acidic residues in CaM contribute to determining the nature of the lowest energy conformation in solution and that electrostatic interactions between these residues are dependent on pH [48]. In fact, the linker has been studied as an intrinsic property and 3 ns long MD simulations were conducted on the linker region. Bending motions operating on 0.5 ns time scale were observed [94]. The coupling between

bending and torsional modes of the linker was further demonstrated by Monte Carlo methods [95]. Moreover, it was hypothesized that pH- and IS dependent shifts in the populations of conformational substates result from changes in electrostatic interactions in the central linker [47, 48]. For example, the shift in favor of the more compact conformation at reduced pH may result from the loss of electrostatic interactions that serve as spacers at neutral pH. This hypothesis is corroborated by inspection of the proximity of side chains of glutamic and aspartic acid residues surrounding the hinge region in the compact Ca^{2+} -CaM crystal structure [27].

Also, the dynamical aspects of Ca^{2+} removal from NTD was studied and it resulted in compaction of the extended initial structure within 7.5 ns in a 12.7 ns trajectory where the trajectory of a fully Ca^{2+} loaded CaM resulted in no conformational change [96]. Further, a coupled MD - MM approach leads to the fast propagation of the intradomain conformational changes obtained upon removal of all Ca^{2+} ions where the method was applied separately to either NTD or CTD [97] .

Thus, independent of the water model, force field type or IS values used, all of the mentioned studies point to the stability of the extended form Ca^{2+} -CaM, despite the high degree of flexibility observed in the linker. In this study, we couple extensive MD simulations with the methods mentioned previously to thoroughly investigate the conformational multiplicity of Ca^{2+} -CaM.

1.3. Motivation

In this study we aim to provide an explanation for the significantly large conformational changes occurring in Ca^{2+} -CaM in response to small perturbations that may arrive at local regions by using coarse-grained (PRS) and all-atom (MM, MD) methods. We aim to decipher the key residues in Ca^{2+} -CaM that may be targeted to facilitate conformational change when exposed to a large number of possible perturbations existing in nature randomly.

While studying the dynamics of Ca^{2+} -CaM under different perturbation scenarios, we also strive to determine design parameters generalizable to other proteins displaying conformational multiplicity. Another motivation is to pinpoint how shifts in the energy landscapes may be induced. We also aim to determine optimal methods as well as deciphering the differences between applying external perturbations in the form of forces vs. displacements.

CHAPTER 2

Theoretical and Computational Methods

2.1. Protein Study Set

CaM is a small acidic protein of 148 residues. In studies of Ca^{2+} -CaM, the extended conformation with a dumbbell shape containing two domains joined by an extended linker is customarily used (Figure 2). CaM contains four EF hand motifs. This motif consists of a simple helix-loop-helix architecture in which the interhelical loop region contains several amino acids essential to the coordination of a single Ca^{2+} ion [20]. There are two globular structural units in CaM, termed N terminal (NTD) and C terminal domains (CTD), including residues 5-68 and 92-147, respectively. The linker is defined by residues 69-91. It should be mentioned that the domain boundaries are not crucial for the aims of the present study and similar boundaries were obtained using a variety of domain parsers [98]. In addition, residues 1-3 are disordered in some of the protein data bank structures used in this study as well as residue 4 missing side chain atoms. The last residue, 148, is also missing in some X-Ray structures. Thus, the first four and last one residue is deleted from these structures and the lobe boundaries are defined accordingly.

All CaM structures studied in this work contain four Ca^{2+} ions, two bound to each domain. EF hand structure in each domain is a 12-residue-long motif, whereby positions 1-3-5 are occupied by Asp or Asn residues which act as monodentate Ca^{2+} ligands, and position 12 is occupied by a bidentate Asp or Glu residue. For the CaM structures studied here, the coordinating residues in each of the four EF-hands are as follows: loop I (D20-D22-D24-E31), loop II (D56-D58-N60-E67), loop III (D93-D95-N97-E104), loop IV (D129-D131-D133-E140).

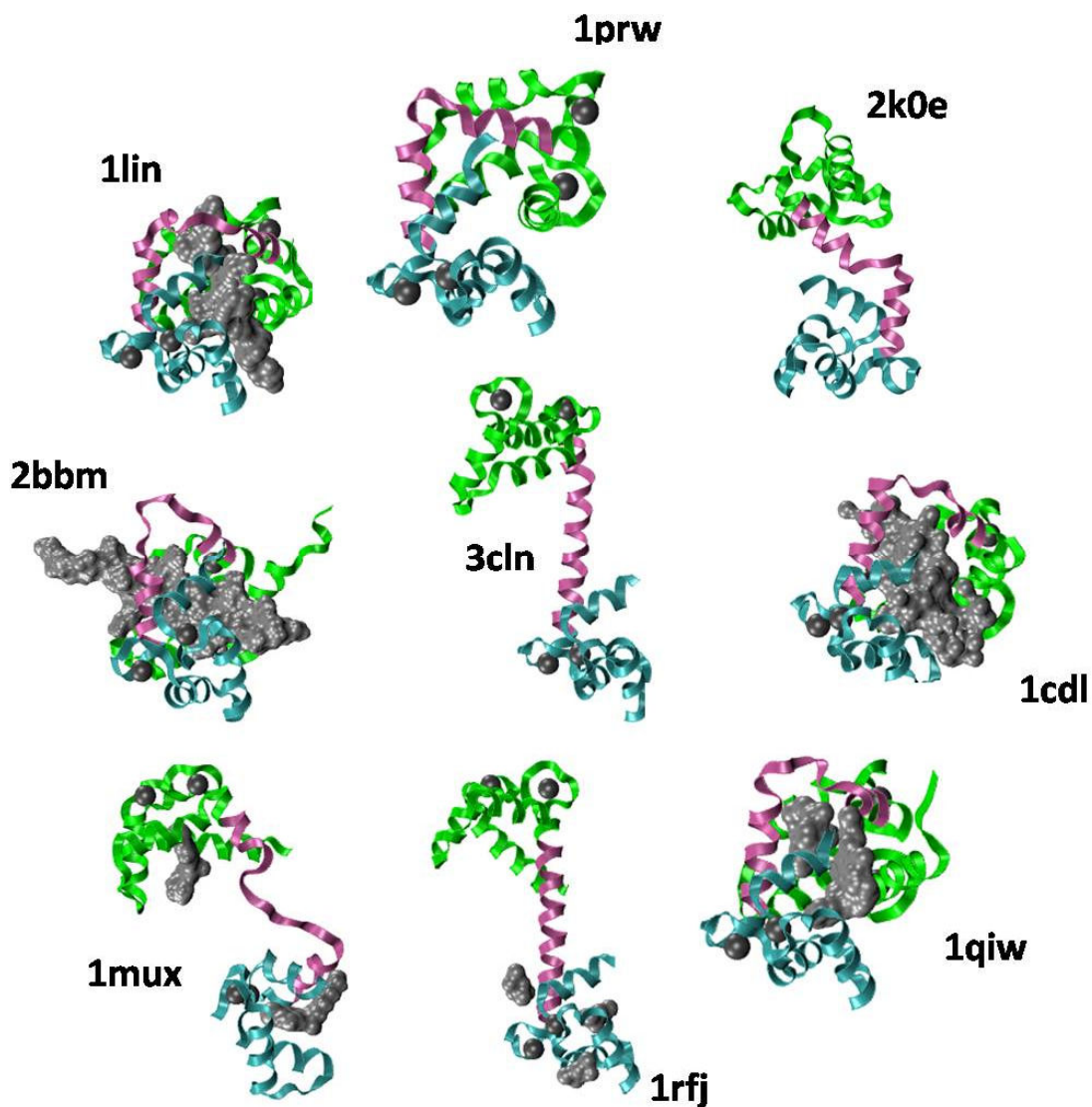


Figure 2. Three-dimensional structures of the proteins studied in this work. The initial structure is placed in the center, and the targets are oriented in such a way that their C-terminal domains are best fitted. N-terminal domain is in green, C-terminal domain is in cyan, and the linker is in magenta. The Ca²⁺ ions are shown as gray spheres, the bound ligand molecules are shown in gray surface representations.

Here we study nine CaM structures in which three of them are unliganded (apo) forms and the six of them are liganded (holo) forms. The bound ligands, the PDB codes, the experimental resolution for those structures determined by X-ray methods, and the source organisms as well as the sequence identity to the extended apo CaM structure are listed in Table 1.

Table 1. Calmodulin structures studied and their properties.

PDB ID	Source	Sequence Identity ^a	Experimental Resolution (Å)	Ligand
1lin	<i>Cow (Bos taurus)</i>	99.3	2.0	Four trifluoperazine (TFP) groups, each having 23 heavy atoms
1qiw	<i>Cow (Bos taurus)</i>	98.6	2.3	Two diphenylpropyl-bis-butoxyphenyl ethyl-propylene-diamine (DPD) groups each having 28 atoms
2bbm	<i>Drosophila melanogaster</i>	93.9	NMR	Myosin light chain kinase (26 residues)
1cdl	<i>Homo sapiens</i>	98.6	2.0	Protein kinase type II alpha chain (19 residues)
1mux	<i>African frog (Xenopus laevis)</i>	99.3	NMR	Two aminohexyl-chloro-naphthalenesulfonamide (WW7) groups, each having 22 heavy atoms
1rfj	<i>Potato (Solanum tuberosum)</i>	88.8	2.0	Three methyl-pentanediol (MPD) groups each having 8 heavy atoms
1prw	<i>Cow (Bos taurus)</i>	99.3	1.7	No ligand
2k0e	<i>Homo sapiens</i>	95.9	NMR	No ligand
3cln	<i>Ratus ratus</i>	-	2.2	No ligand

^aTo the extended structure, PDB code 3cln.

2.2. System Construction and Molecular Dynamics Simulation Protocol

We have performed MD simulation using 3cln, 1prw and 2k0e as the initial structures of CaM and prepared different systems. The properties of the systems used in the MD simulations are given in Table 2. Using VMD 1.8.7 program with solvate plugin version 1.2 all systems are solvated such that at least 10 Å from each direction is water. [99]. The NAMD package is used to model the dynamics of protein-water system [100]. The CharmM27 force field parameters are used for protein and water molecules [65]. Water molecules are described by the TIP3P model. The initial box dimensions, the number of atoms each system contains, the corresponding ionic strengths of each system are listed in Table 3. Long range electrostatic interactions are calculated by the particle mesh Ewald method [101] with a cutoff distance of 12 Å and a switching

distance at 10 Å. RATTLE algorithm [102] is applied to use a step size of 2 fs in the Verlet algorithm [101]. Temperature control is carried out by Langevin dynamics with a dampening coefficient of 5/ps. Pressure control is attained by a Langevin piston with Langevin period of 100 fs and Langevin decay of 50 fs. Volumetric fluctuations are preset to be isotropic. The system is run in the NPT ensemble at 1 atm and 310 K until volumetric fluctuations are stable to maintain the desired average pressure. In this case, this process requires a 2 ns long equilibration period. The respective box dimensions of the equilibrated structures are also listed in Table 3. The run in the NPT ensemble is extended to a total of at least 100 ns for each system. The coordinate sets are saved at 2 ps intervals for subsequent analysis, leading to at least $T = 50\,000$ snapshots. For several systems independent runs are performed to check the reproducibility of the results (Table 2).

Table 2. Features of the protein-water systems used in the MD simulations.

Initial Conformation	Conditions*	IS (mM)	Atom #	Equilibrated Box size (Å)	Simulation length (ns)	Run Label
Extended (3cln)	Residues standard pK _a s @ pH 7.4	82	37099	62x87x69	400	3CLN
Compact (1prw)	Residues standard pK _a s @ pH 7.4	161	20330	58x57x60	200 200	1PRW
Extended (3cln)	Residue 31 is protonated.	91	31652	63x81x60	150 400	E31 ⁺
Extended (3cln)	Residue 31 is mutated to Alanine	94	31645	60x83x62	400 400 400	E31A
Extended (3cln)	Residues 11, 31, 67, 84, 93, 104, 107, 122, 133, 140 are protonated.	43	31660	64x81x60	100 200	3CLN*
Extended (3cln)	Residues standard pK _a s @ pH 7.4	150	33835	82x84x90	150	3CLN ^{150mM}

*in all runs temperature is 310 K and pressure is 1 atm, differences in IS and protonation are listed.

The correlations between residue pairs derived from the MD trajectory are of particular interest. We consider the Cartesian coordinates of the C_α atoms and Ca²⁺ ions recorded at each time step t in the form of the $3N \times 1$ coordinate matrix, $\mathbf{R}(t)$, where N is the number of nodes. This matrix does not contain the original coordinates from the MD

trajectories, but instead, it is obtained by the best superposition of all the T structures. A mean structure $\langle \mathbf{R}_i(t) \rangle$ is defined as the average over these coordinate matrices. One can then write the positional deviations for each residue i as a function of time and temperature, $\Delta \mathbf{R}_i(t) = \mathbf{R}_i(t) - \langle \mathbf{R}_i(t) \rangle$. These are organized as the columns of the $3N \times T$ fluctuation trajectory matrix, $\Delta \mathbf{R}$,

$$\Delta \mathbf{R} = \begin{bmatrix} \Delta \mathbf{R}_1(t_1) & \Delta \mathbf{R}_1(t_2) & \cdot & \Delta \mathbf{R}_1(t_T) \\ \Delta \mathbf{R}_2(t_1) & \Delta \mathbf{R}_2(t_2) & \cdot & \Delta \mathbf{R}_2(t_T) \\ \Delta \mathbf{R}_3(t_1) & \Delta \mathbf{R}_3(t_2) & \cdot & \Delta \mathbf{R}_3(t_T) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \Delta \mathbf{R}_N(t_1) & \Delta \mathbf{R}_N(t_2) & \cdot & \Delta \mathbf{R}_N(t_T) \end{bmatrix} \quad (1)$$

2.3. Local Perturbations Introduced by Perturbation Response Scanning (PRS)

Linear response theory (LRT) has been recently used to study conformational changes undergone by proteins under selected external perturbations [36]. In this study, we use a toolkit that we term perturbation-response scanning (PRS) which is based on sequential application of LRT to study the origins of structural changes undergone by protein molecules [22].

PRS relies on systematically applying forces at singly selected residues and recording the linear response of the whole protein. The response is quantified as both the magnitude of the displacements undergone by the residues, and their directionality. Since the procedure is repeated sequentially for all the residues in the protein, we term the technique, perturbation-response scanning (PRS). Below, we first review the theory and then outline the details of the PRS technique.

2.3.1. Linear Response Theory

The derivation of how a structure may be manipulated by external forces is presented in below [76, 103]. We construct the protein as a residue network of N nodes that are centered on the C_α atoms. Any given pair of nodes are assumed to interact via a harmonic potential, if they are within a cut-off distance r_c of each other.

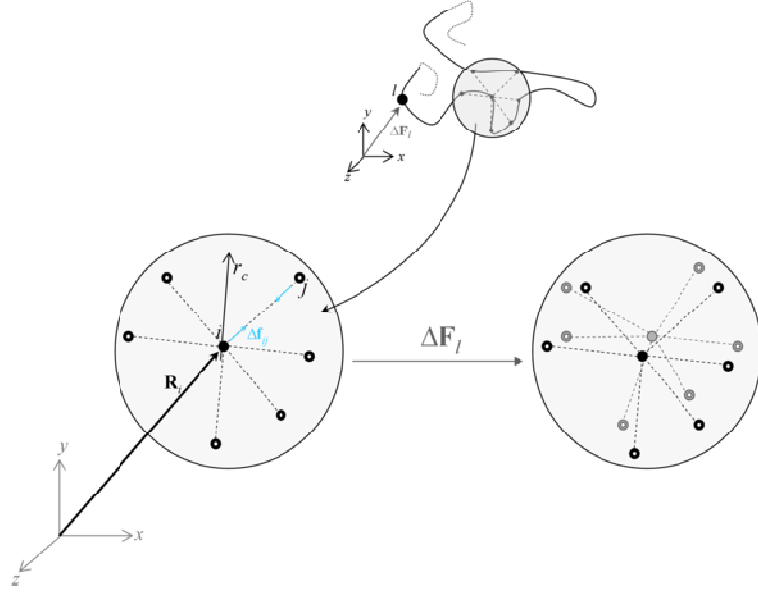


Figure 3. Free-body diagram representation of a residue.

In the notation used, \mathbf{r} and \mathbf{f} refer to the bond and internal force vectors along the edge connecting any two nodes, respectively. On the other hand, \mathbf{R} and \mathbf{F} are vectors on the nodes and are referred to as the position and external force vectors, respectively. There are m interactions pertaining to each residue (Figure 3, as an example, schematically illustrates the interactions for a residue that has six interactions, *i.e.*, $m = 6$), and a total of M interactions for the system of N residues. In the absence of an external force acting on the system, the equilibrium condition for each residue, i , necessitates that the summation of the internal, residue-residue interaction forces must be zero for each residue. Therefore,

$$\mathbf{b}\Delta\mathbf{f}_i = 0 \quad (2)$$

where the $3 \times m$ coefficient matrix \mathbf{b} consists of the direction cosines of each force representing the residue-residue interaction. The row indices of \mathbf{b} are x, y , or z . Here $\Delta \mathbf{f}_i$ is an $m \times 1$ column matrix of forces aligned in the direction of the bond between the two interacting residues. For instance, in Figure 3, residue i has six contacts; and, thus, $\Delta \mathbf{f}_i$ is a 6×1 column matrix. Following the example outlined in Figure 3, equation 2 sum up the projection of these six forces on the x, y , and z -axes. This algebra gives rise to three independent equations involving six unknown interaction forces, which are the residual interaction forces of residue i with its contacting neighbors. One can write the equilibrium condition (equation 2) for each residue. This results in a total of N sets of equations, each of which involves the summation of forces in three respective

directions. Consequently, generalizing equation 1 to the whole system of N nodes and a total of M interactions, one can write the following algebraic system of a total of $3N$ number of equations consisting of M number of unknown residue-residue interaction forces

$$\mathbf{B}\Delta\mathbf{f}_i = 0 \quad (3)$$

with the $3N \times M$ direction cosine matrix \mathbf{B} and the $M \times 1$ column matrix of residue-residue interaction forces, $\Delta\mathbf{f}$. It is straightforward to generate the matrix \mathbf{B} from the topology of the native structure (*i.e.*, the protein data bank (PDB [68]) file) with a specified r_c .

In the presence of an external force, $\Delta\mathbf{F}$ (Figure 3), the equilibrium consideration for each residue dictates that the summation of the residue-residue interaction forces must be equal to the external, applied force on the same residue. Then, equation 3 may be cast into the following form

$$\mathbf{B}_{3N \times M} \Delta\mathbf{f}_{M \times 1} = \Delta\mathbf{F}_{3N \times 1} \quad (4)$$

Under the action of external forces, each residue experiences a displacement, $\Delta\mathbf{R}$, which is termed the positional displacement vector. Moreover, the bond distance between any two residues changes in the amount of $\Delta\mathbf{r}$ in accord with the positional displacements of the two residues which participate in the bonding. Therefore, there must be compatibility between the total of $3N$ number of positional displacements and the changes that take place in the intra-residual distances, a total of M number of distortions. This compatibility is very similar to the form given in equation 4 [103]:

$$\mathbf{B}_{M \times 3N}^T \Delta\mathbf{R}_{3N \times 1} = \Delta\mathbf{r}_{M \times 1} \quad (5)$$

Within the scope of an elastic network of residues that are connected to their neighbors with linear-elastic springs, the residual interaction forces, $\Delta\mathbf{f}$, are related to the changes in the contact distances, $\Delta\mathbf{r}$, through Hooke's law by

$$\mathbf{K}_{M \times M} \Delta\mathbf{r}_{M \times 1} = \Delta\mathbf{f}_{M \times 1} \quad (6)$$

where the coefficient matrix \mathbf{K} is diagonal. Thus, rearranging equations 4–6, one gets the forces necessary to induce a given point-by-point displacement of residues:

$$(\mathbf{BKB}^T)\Delta\mathbf{R} = \Delta\mathbf{F} \quad (7)$$

On the other hand, one may choose to perturb a single or a set of residues, and follow the response of the residue network through,

$$(\mathbf{BKB}^T)^{-1}\Delta\mathbf{F} = \Delta\mathbf{R} \quad (8)$$

where the $\Delta\mathbf{F}$ vector will contain the components of the externally applied force vectors on the selected residues. The (\mathbf{BKB}^T) matrix is equivalent to the Hessian [76] and its inverse has six zero eigenvalues, corresponding to the global translational and rotational degrees of freedom of the system. The elements of the inverse of the Hessian, $\mathbf{G} = \mathbf{H}^{-1}$, may be used to predict the auto- and cross-correlations of residues. \mathbf{G} may be viewed as an $N \times N$ matrix whose ij th element is the 3×3 matrix of correlations between the x -, y -, and z -components of the fluctuations $\Delta\mathbf{R}_i$ and $\Delta\mathbf{R}_j$ of residues i and j ; *i.e.*,

$$\mathbf{G}^{ij} = \begin{bmatrix} \langle \Delta X_i \Delta X_j \rangle & \langle \Delta X_i \Delta Y_j \rangle & \langle \Delta X_i \Delta Z_j \rangle \\ \langle \Delta Y_i \Delta X_j \rangle & \langle \Delta Y_i \Delta Y_j \rangle & \langle \Delta Y_i \Delta Z_j \rangle \\ \langle \Delta Z_i \Delta X_j \rangle & \langle \Delta Z_i \Delta Y_j \rangle & \langle \Delta Z_i \Delta Z_j \rangle \end{bmatrix} \quad (9)$$

The cross-correlations between residue pairs are obtained from the trace of its components:

$$\langle \Delta\mathbf{R}_i \Delta\mathbf{R}_j \rangle = \text{tr}(\mathbf{G}^{ij}) \quad (10)$$

Equation 10 has been shown to reproduce the cross-correlations obtained from MD simulations and MM [35, 87]. In this work, we shall not be directly interested in the correlations, but rather shall use \mathbf{G} as a kernel to predict the response of other residues to applied perturbations on selected ones as we discuss next.

Thus, the $3N \times 3N$ correlation (or covariance) matrix is calculated from equation 1 by $\mathbf{G} = \Delta\mathbf{R}\Delta\mathbf{R}^T$, where T is the transpose.

2.3.2. Perturbation-Response Scanning (PRS) Protocol

Our detailed PRS analysis is based on a systematic application of equation 7. We apply a force on the C_α atom of each residue by forming the $\Delta\mathbf{F}$ vector in such a way that all the entries, except those corresponding to the residue being perturbed, are equal to zero. For a selected residue i , the external force vector is constructed as,

$$(\Delta\mathbf{F})^T = \{000 \dots \Delta F_x^i \Delta F_y^i \Delta F_z^i \dots 000\} \quad (11)$$

The direction of the applied force vector deserves special attention. Here we choose the forcing direction randomly, attributing no bias due to the specific contact topology or the solvent exposed nature of the residue being perturbed. The forcing directions are uniformly distributed within a sphere enveloping the residue; therefore, the forcing may well be termed isotropic. It is definitely possible to favor specific directions leading to anisotropy in forcing, since there are no intrinsic constraints in the methodology dictating the opposite. A plausible forcing scenario for contact with a ligand, similar to the ones in [36] may also be conceived to determine the associated conformational changes. We then compute the resulting $(\Delta\mathbf{R})$ vector of the protein through equation 8.

2.4. Local Perturbations Introduced by Molecular Mechanics

Molecular mechanics (MM) is the application of classical mechanics to molecules. Classical mechanics is used to describe the motion of macroscopic objects. In molecular mechanics, atoms are treated as spheres whose mass depends on the element. Chemical bonds are treated as springs whose stiffness depends on which elements are bound together, and whether the bond is single, double, or triple. Other types of springs are used to model changes in bond angles, dihedral angles, etc. Each of these various types of springs will have spring constants associated with them. Additional equations from classical physics, such as Coulomb's Law, are used to handle any electrostatic interactions present within a molecule. The sum of all energy terms that apply to a particular molecule are added together to define what is called the "steric" energy, or total potential energy, of the molecule. All of the equations and associated parameters used to calculate each energy term are collectively called the "force field". The basic application of molecular mechanics is the energy minimization.

In the present study, we will use MM in line with perturbations in the form of selected atomic displacements.

This is a simple approach whereby the energy minimized native coordinates of the protein is altered by inserting a displacement at a selected residue followed by energy minimization. When a small perturbation is introduced in a region of the molecule, relaxation to equilibrium does not occur by directly restoring to the initial conditions [104, 105]. For example, if a bond is lengthened, this bond will not directly shrink to its original coordinates during energy minimization. Rather, relaxation at softer degrees of freedom will take precedence, and rearrangements, e.g., in the torsional degrees of freedom, will also be observed. Therefore, it is assumed that the perturbation is applied quasi-statically. Such a disturbance is related to mutations, i.e. alterations of packing topology. This method is also used to show that residue stability arises as a tool reflecting the character of the response [35].

In this method, the interest is given to the relative displacements of the atoms, and the criterion for terminating the minimization must be selected accordingly. In a previous study it was shown that the usage of a more stringent criterion (e.g., 10^{-4} kcal/mol/Å) affected the absolute value of the energy attained by ~ 0.2 kcal/mol, whereas the final atomic locations were within ~ 0.01 Å of their absolute minima at a derivative of 0.1 kcal/mol/Å [35]. It was also shown that this was especially true when the minimization was started from a slightly distorted structure obtained from a very well minimized one as it will done in this work. Therefore, a termination criterion of 0.1 kcal/mol/Å of the derivative in all the subsequent energy minimizations will be employed.

The CharmM27 [65] force field in conjunction with the NAMD software package is used in the MM calculations [100]. Initially, Ca^{2+} -CaM in extended state (3cln), snapshot taken from the well equilibrated 20 ns point of the MD trajectory is energy minimized for 20000 steps with conjugate gradients algorithm. Group-based cutoffs are employed with a 12 Å cutoff distance. A switching function is used at 10 Å. The RMSD between the original PDB and the energy minimized C_α coordinates is 3.5 Å. These coordinates of the molecule are treated as the main coordinates of the equilibrium structure.

We use a procedure of systematically perturbing each of the protein residues according to the following scheme: (1) The coordinates of the C_α of the selected residue are displaced by an amount d_i , which leads to a distorted local structure; (2) the new

coordinates of the displaced C_α atom is fixed in space, whereas the rest of the protein atoms are free to move; (3) the energy of the protein is minimized to 0.1 kcal/mol/Å of the derivative; (4) the rearranged atomic coordinates in response to the perturbation are recorded for further analysis; (5) the coordinates are reset to those of the equilibrium structure and the process (1–4) is repeated over all the C_α atoms. Note that the perturbation of residue i mentioned in item (1) may be carried out in several different ways. Here, the displacement perturbation is given to a residue such that each time the residue will be perturbed in a surface of a sphere that has a radius of 0.5 Å. This is simply done by, (i) first generating a random displacement vector in (x,y,z) directions, (ii) calculating the magnitude of this displacement vector (m), (iii) multiplying the displacement vector by $(R/m)^2$, where R is the radius of the sphere.

The N sets of protein coordinates obtained at the end of the MM calculations are organized into the perturbation-response matrix of order $3N \times 3N$, where the displacement vector of residue i in response to a directed perturbation placed at residue k is denoted by $\Delta \mathbf{R}_{ik}$:

$$\Delta \mathbf{R} = \begin{bmatrix} \Delta \mathbf{R}_{11} & \Delta \mathbf{R}_{12} & \cdots & \Delta \mathbf{R}_{1N} \\ \Delta \mathbf{R}_{21} & \Delta \mathbf{R}_{22} & \cdots & \Delta \mathbf{R}_{2N} \\ \vdots & \vdots & & \vdots \\ \Delta \mathbf{R}_{N1} & \Delta \mathbf{R}_{N2} & \cdots & \Delta \mathbf{R}_{NN} \end{bmatrix} \quad (12)$$

Then, the $N \times N$ displacement matrix, $\Delta \mathbf{L}$, each of whose elements, $\mathbf{L}_{ik} (= \Delta \mathbf{R}_{ik})$, are the magnitudes of $\Delta \mathbf{R}_{ik}$ and hence is a measure of the amount of displacement experienced by i in response to the perturbation placed at k :

$$\mathbf{L} = \begin{bmatrix} \Delta \mathbf{R}_{11} & \Delta \mathbf{R}_{12} & \cdots & \Delta \mathbf{R}_{1N} \\ \Delta \mathbf{R}_{21} & \Delta \mathbf{R}_{22} & \cdots & \Delta \mathbf{R}_{2N} \\ \vdots & \vdots & & \vdots \\ \Delta \mathbf{R}_{N1} & \Delta \mathbf{R}_{N2} & \cdots & \Delta \mathbf{R}_{NN} \end{bmatrix} \quad (13)$$

Cross correlations between residues i and j in response to the inserted perturbations is given by Eq. 10 where $\mathbf{G} = \Delta \mathbf{R} \Delta \mathbf{R}^T$.

2.5. Steered Molecular Dynamics Simulations

Steered molecular dynamics (SMD) is a common method that is used to explore the biological processes on time scales accessible to MD simulations. For instance, unbinding of ligands, and conformational changes in biomolecules may be studied using this technique. SMD is an extended MD simulation that mimics the atomic force microscopy (AFM) experiments

The basic idea behind any SMD simulation is to apply an external force to one or more atoms. Proteins can be stretched or unfolded by pulling one atom using this technique. This technique can be applied in two ways either pulling with a constant velocity or constant force. In constant velocity pulling, the atom to be pulled is virtually attached to a dummy atom. Then the dummy atom is moved with a constant velocity and the force between the atom of interest and dummy atom is measured. The following equation is valid for this scenario:

$$\vec{F}(t) = -\nabla \left(k \left[vt - (\vec{r} - \vec{r}_0) \cdot \vec{n} \right] \right) \quad (14)$$

where k is the spring constant, v is the pulling velocity, t is the time, \vec{r} and \vec{r}_0 are the actual and initial position of the SMD atom, respectively. As the dummy atom moves with a constant velocity, the SMD atom experiences a force that depends linearly on the distance between the dummy and the SMD atom as shown in Figure 4.

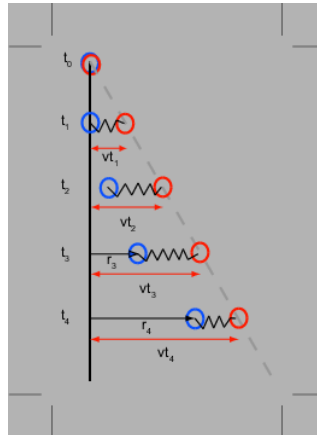


Figure 4. Pulling in one dimensional case. Dummy atom and SMD atom are colored in red and blue, respectively. Adapted from [106].

The key parameters for this technique are the pulling velocity and the spring constant. These values may be determined directly from experiments (AFM experiments for the system of interest) or by running several trials.

The second way to carry out a SMD simulation is by application of constant force. In this case the SMD atom experiences a constant force in the direction of the vector that links the fixed and the pulled atoms. In this case there is no dummy atom or virtual spring.

SMD allows exploring the binding and unbinding properties of a variety of biomolecules as well as their responses to external mechanical manipulations at the atomic level. In this work constant force pulling is used. Traditional analysis that is used in literature will not be applied. The direction of the force is received from the PRS methodology and applied to the residues that are implicated by PRS. After steering the residue in the specific direction the displacements of the whole system are investigated.

2.6. Quantifying Structural Similarity by Overlap

At the end of the MM and PRS calculations, we measure the overlap to quantify similarity of the displacements coefficient.

$$O^{jk} = \frac{\Delta \mathbf{S}^j \cdot \Delta \mathbf{S}^k}{|(\Delta \mathbf{S}^j \cdot \Delta \mathbf{S}^j)(\Delta \mathbf{S}^k \cdot \Delta \mathbf{S}^k)|^{1/2}} \quad (15)$$

Here, the superscripts j and k refer to different three-dimensional structures and $\Delta \mathbf{S}^j$ and $\Delta \mathbf{S}^k$ are the displacement vectors between every protein pair listed in Table 4 and 3cln-newcoordinates upon perturbation respectively. O^{jk} is a measure of the similarity of directionality of the conformational change that occurs upon binding. While $O^{jk} = 1$ represents a perfect overlap of the directionality of the conformational change, the root mean square deviation (RMSD) of two structures that have such an overlap value need not be zero. Two structures maybe moving along the same vector, but if the amount of the move is varied, they would yield different RMSD values. For example, consider the simple hinge motion of three points in space, where the closing motion may have proceeded by either a small or a large amount. The RMSD between these two configurations would be large compared to their overlap, measured as the angle between the two lines of motion (also see Fig 1 in ref [75]). Thus, RMSD and overlap yield complementary information. Note that how structural alignment is carried

out affects the overlap values; finding regions within the protein that give higher overlap value is possible. We chose to align over the whole protein to make sure that we do not bias some of the conformational changes.

2.7. Reduced Degrees of Freedom of CaM

A simplified set of coordinates are used to capture the main features of the relative motions of the NTD, CTD and linker region of CaM by reducing the structure to four points in space. These points are schematically shown in Fig. 5. Two of these points are the center of masses (COMs) of the NTD (point 1), CTD (point 4). In addition, residues 69 and 91 are used to mark the beginning and end points of the linker (points 2 and 3). We then define two degrees of freedom to capture the essence of the motions of the two lobes relative to each other. Torsion angle φ defines the rotation of the two lobes relative to each other while linker end-to-end distance l_{linker} defines spatial orientation of the lobes. All the structures used in this study are represented in the l_{linker} vs φ plane and shown in Figure 5.

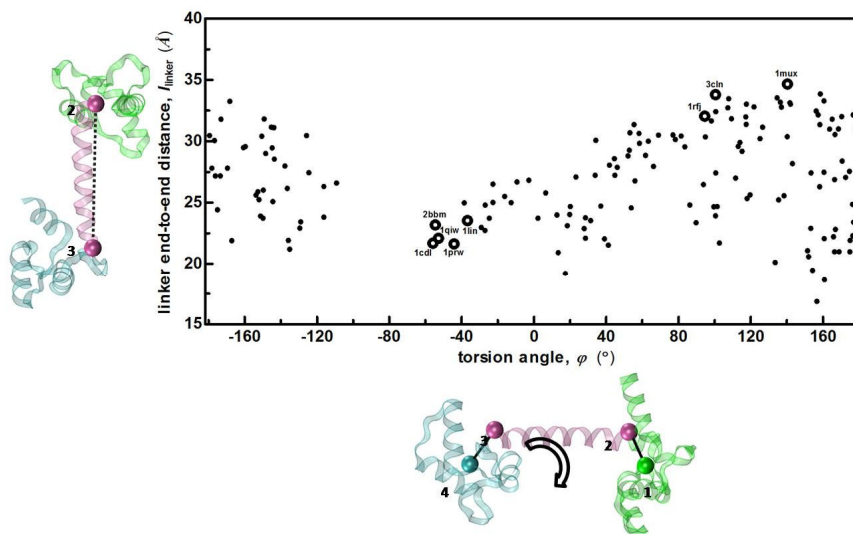
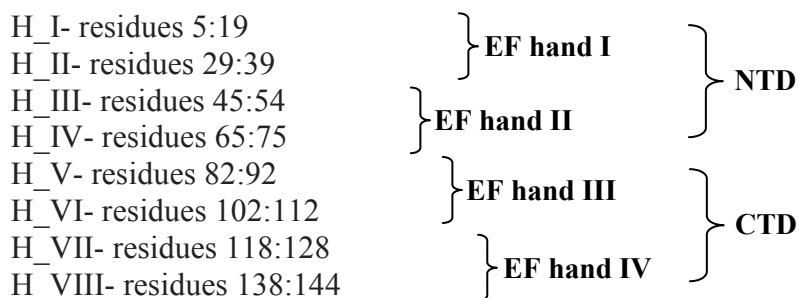


Figure 5. l_{linker} vs φ plot representing CaM structures in the study set. Black full circles represent the 2k0e NMR ensemble. Open circles are the X-ray structures in the study set.

2.8. Inter-Helical Angles

In order to monitor the changes in the dynamics of the EF hand motifs in each domain, the angles between the helices in each EF hand is calculated as a trajectory for each run. The helices are numbered as follows:



For each snapshot taken from the trajectory, C_α coordinates are organized in the matrix of $3 \times H$, where H is the number of residues that constitutes the helix. Prior to the matrix formation, to remove the effect of rotation and translation, the center of mass for each helix is calculated and it is subtracted from the original coordinates and modified coordinates are used in the matrix formation. Then, for each matrix representing the helices, singular value decomposition is performed and the vector that corresponds to the highest singular value is accepted as the main axis of that helix. The angle between these vectors is then calculated from the dot product of these vectors.

2.9. pKa Calculations

For the pK_a and degree of ionization calculations, we mainly used the pH-dependent Protein Electrostatics Server (PHEPS) program implemented in the PHEMTO server [107, 108]. The method is based on a self-consistent approach to calculating protein electrostatics. The intrinsic pK_a value is defined as the modification of the pK_a in the model compounds by the Born energy and the contributions from the partial charges of interacting atoms. Starting from a set of initial values, the electrostatic free energy is calculated iteratively until the pK_a values converge. The effect of the bound Ca^{2+} ions is also included in the calculations. We also perform pK_a calculations using the PROPKA[109], H^{++} [110], and pKD [111] servers. PROPKA and pKD also rely on the accurate calculation of the shifts in free energies. The latter particularly focuses on a correct representation of hydrogen-bonding interactions, while the former is based on an improved description of the desolvation and the dielectric response of the

protein. We note that while PROPKA version 3.0 has an improved representation of the titration behavior, it does not yet include ions explicitly in the calculations; we therefore used version 2.0 in the calculations. The H^{++} server uses a different approach than direct calculation of free energy changes, whereby the complicated titration curves are directly represented as a weighted sum of Henderson-Hasselbalch curves of decoupled quasi-sites. We report values for the settings of 150 mM salinity, external dielectric of 80 and internal dielectric of 10. The latter is a suggested value for better prediction of solvent exposed residues, although we have checked that the trends are not affected by this choice. The calculated pK_a values of CaM using PDB structure 3cln by these four methods are provided in Chapter 3 section 3.2. Where possible, experimentally measured values from literature are also included [112, 113], along with the reported calculations on the PDB structure 1cll, another x-ray structure for Ca^{2+} -CaM using the multi-conformation continuum electrostatics (MCCE) method [114].

CHAPTER 3

Results

3.1. Structure and Sequence Analysis of the Proteins

For every pair in the study set listed in Table 1, Structural Alignment of Multiple Proteins (STAMP structural alignment), implemented in VMD 1.8.7 MultiSeq plugin was performed [115]. The root mean square deviation (RMSD) between each pair in the study set was calculated (Table 3). Also, RMSD of the NTD and CTD values were calculated separately for each protein pair. It was found that the overall RMSD between the extended (3cln) and other structures are mostly on the order of 15-16 Å, except for 1rfj (2.7 Å) and 1mux (6.4 Å). For the ligand bound structures, the magnitude of the change do not depend on the ligand size or on the region of the protein it binds. In contrast, the superposition of only NTD or CTD yield low RMSD; the only ones that have values above 1.5 Å are the NTD of 1prw and 2bbm. This result is the implication of the similar internal arrangements in the two lobes and hints that the conformational change mostly involves global motions rather than local rearrangements.

Also, the RMSD amongst the other eight structures themselves are compared to quantify the amount of structural difference they have. 1rfj and 1mux both had RMSD in the interval of 14-16 Å with the other six structures. It should be noted that since these two structures are also the ones that have lower RMSD with 3cln, they might be located closer to 3cln in the conformational space (CS), and at a part different from the other six structures. They are not, however, in exactly the same region of the CS since the overall RMSD between them is large (7.4 Å). The internal arrangement of both NTD and CTD are similar with 1.3 and 1.1 Å RMSD, respectively; so the structural difference must be in their relative positioning. Also, displayed in Table 3 are the overlaps between the displacements vectors of the experimental structures (Eq.15).

Table 3. RMSD between pairs of structures listed in Table I. Lower diagonal: RMSD between overall structures; upper diagonal: RMSD between NTD (bold) and CTD (italic) only. Overlap between the experimental displacement vectors ΔS to extended conformation are also displayed in parentheses.

PDB ID	3cln	1prw	2k0e*	1lin	1qiw	2bbm	1cdl	1rfj	1mux
3cln		2.3 <i>1.1</i>	1.86 <i>1.52</i>	0.63 <i>0.70</i>	0.39 <i>0.95</i>	1.9 <i>1.5</i>	0.51 <i>0.76</i>	0.63 <i>0.80</i>	1.2 <i>1.3</i>
1prw	16		14.5	2.4 <i>0.70</i>	2.8 <i>0.70</i>	3.3 <i>1.2</i>	2.5 <i>1.0</i>	2.1 <i>1.3</i>	2.5 <i>1.5</i>
2k0e*	7.9	1.45 <i>2.27</i>		1.54 <i>1.97</i>	1.85 <i>1.68</i>	2.73 <i>2.66</i>	1.63 <i>1.41</i>	1.56 <i>1.35</i>	1.68 <i>2.08</i>
1lin	15	4.2 (0.96)	14.5		0.68 <i>0.70</i>	2.1 <i>1.2</i>	0.51 <i>0.80</i>	0.45 <i>0.91</i>	1.1 <i>1.3</i>
1qiw	15	3.6 (0.95)	15	1.6 (0.99)		1.8 <i>1.4</i>	0.57 <i>0.90</i>	0.68 <i>0.80</i>	1.2 <i>1.1</i>
2bbm	15	5.2 (0.89)	16	3.5 (0.94)	2.5 (0.96)		1.9 <i>1.4</i>	2.0 <i>1.7</i>	2.1 <i>1.4</i>
1cdl	15	4.6 (0.91)	15.3	2.8 (0.97)	1.9 (0.99)	2.2 (0.96)		0.48 <i>0.62</i>	1.2 <i>1.1</i>
1rfj	2.7	16 (0.17)	8.6	15 (0.09)	15 (0.09)	15 (0.05)	15 (0.03)		1.3 1.1
1mux	6.4	16 (0.17)	11.2	15 (0.09)	14 (0.23)	14 (0.26)	14 (0.28)	7.4 (0.20)	

*RMSDs calculated by 1st structure from the NMR ensemble.

Inspection of O^{ik} values listed in Table 3 shows that 1lin, 1prw, 1qiw, 2bbm and 1cdl display similar types of conformational motions ($O^{ik} \geq 0.89$), while 1rfj and 1mux each have distinct conformational changes from the rest (O^{ik} is in the range 0.03-0.28), as well as from each other ($O^{1rfj-1mux} = 0.20$). RMSD and overlap values imply a closing of the two lobes towards each other. Also, despite the variety of the ligand types and ligand sizes in the bound forms, there are three classes of conformational changes.

The most similar pair of structure is 1qiw/1lin, which both have large groups binding in the region between the two lobes. Moreover, the internal structures of the NTD and CTD are almost the same (RMSD is 0.7 Å for both lobes, much below the resolution of the x-ray experiments (Table 1)). The remaining pairs have RMSD in the range of 1.9 -5.2 Å. In some cases, the NTD RMSD is above the experimental resolution, e.g., as high as 3.3 Å for the pair 1prw/2bbm. This is in contrast to the rigidity of the CTD, which has an RMSD of less than 1.5 Å in all cases. The

observation of less mobility in CTD is in accord with the higher affinity of the CTD for Ca^{2+} (at 7 μM Ca^{2+} concentration) as opposed to the NTD (at 300 μM Ca^{2+} concentration) [116].

3.2. Perturbation Response Scanning Analysis

Random forces were sequentially inserted on each residue. For each residue, i , the overlap coefficient (Eq.15), O^i between the response vector $\Delta\mathbf{R}^i$ and the experimental conformational change vector $\Delta\mathbf{S}$. In Table 4, the results of the PRS analysis for one compact unliganded structure and six ligand bound structures are reported, using the covariance matrix calculated from the interval between 30 - 120 ns of the trajectory. The reported values represent the single best overlap obtained. It is observed that for none of the conformational changes that is possible to obtain a high overlap by perturbing a single residue in a randomly chosen direction when 3cln crystal conformation ($3\text{cln}^{\text{crystal}}$) was used as the initial structure.

Table 4. Best overlap values obtained for proteins studied by PRS.

Protein pair	Best PRS Overlap ^a	Residue ^a
3cln/1lin	0.72	30, 31, 69
3cln/1prw	0.69	31, 69
3cln/1qiw	0.72	30, 31, 69
3cln/2bbm	0.70	29, 30, 31, 34
3cln/1cdl	0.73	30, 31, 34
3cln/1rfj	0.67	6
3cln/1mux	0.43	Ca^{2+} in loop I

^a results from 500 independent PRS runs using 3cln crystal structure.

By using $3\text{cln}^{\text{crystal}}$ as the initial structure to PRS, it is possible to mimic the different conformational changes to an overlap of 0.70 ± 0.03 for five of the target forms by acting on residue E31 in a selected direction. Neighboring residues to E31 also yield comparable overlap in some cases; these reside on the edge of one of the EF-hand motif loop I (see section 3.1). For the case of 1mux, perturbing Ca^{2+} ion residing in loop

I significantly improves the overlap to 0.43, although this value is well below those for the other target structures. Thus, the initial extended CaM structure (3cln) may be manipulated from this particular EF hand motif. Residue L69 also appears in manipulating 3CLN towards three of the target structures (Table 4). Moreover, overlap values increase when the equilibrated structure at the 20 ns point of the 3CLN MD trajectory is used as the initial structure in PRS. In this case it is possible to mimic the conformational changes to an overlap of 0.76 ± 0.02 for 1prw, 1lin, 1qiw and 2bbm of the target forms by acting on the same regions as the initial structure.

To identify if coupled conformational manipulation improves the results, two residues at a time are perturbed. First, residue pair 31 and 69 are perturbed together as they are the residues yielding high OL for the case of 3cln^{crystal} used as the initial structure. A total of 5000 random perturbations inserted simultaneously on 31/69 did not lead to any improvement of the overlaps. Also, 500 independent scans of coupled perturbations of E31 with all other residues do not identify any improvement of the overlaps. Moreover, 100 iteration scan of all possible node pairs (i.e., $100 \times 147^2 = 2\,160\,900$ independent pair force insertions) confirm that single node perturbations lead to the maximum overlap results for 3cln^{crystal}.

In order to quantify if position 31 is particularly sensitive to subtle pH variations in the physiologically relevant range, we have calculated the degree of ionization of the charged amino acids by using the PHEMTO server [107, 108]. There are 52 titratable groups in CaM, of which 36 are Asp or Glu, 13 are Lys or Arg, two are Tyr, and one is a His. The variation in the degree of ionization as a function of pH is displayed in Figure 6 for four types of charged amino acids.

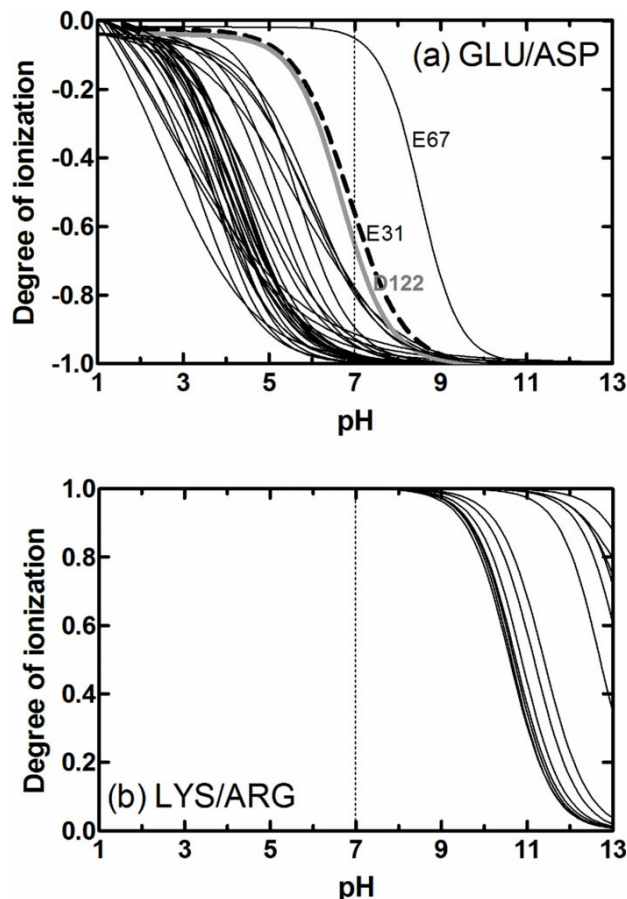


Figure 6. Degree of ionization as a function of pH for (a)Asp/Glu (36 residues) and (b)Lys/Arg (13 residues) amino acid types. E31 (dashed) and D122 (gray) are distinguished as those capable of changing ionization with subtle pH variations at physiological conditions. The standard pKa values for non-perturbed residues are 4.4 for Asp/Glu, 10.0 for Lys, 12.0 for Arg [1].

We find that, only two residues E31 and D122 have large variations in the range of physiologically relevant pH values. The upshift of E31 from the standard value of 4.4 is confirmed by all three other methods (PROPKA[109], H^{++} [110] and pKD[111]) as well as the experimental value reported from the structural homolog of calbindin and MCCE calculations [114] (see Table 5 below). We therefore propose that subtle changes in the pH of physiological environments may be utilized by the protonation/deprotonation of E31, whereby a local conformational change may be translated into the displacement profiles exemplified in Figure. 6, thereby leading to shifts in the conformational energy landscape.

Table 5. pKa values calculated for the structure 3cln by various methodologies; results from other computational (MCCE) and, where possible, experimental work is also included.

residue	H++	propKa 2.0	pKd	PHEMTO	average	standard deviation	MCCE ^a	experimental ^b
GLU 6	4.1	4.6	4.0	3.9	4.1	0.3		
GLU 7	4.2	4.6	4.1	4.1	4.3	0.2	4.7	
GLU 11	5.8	4.8	6.0	5.3	5.5	0.5	5.1	3.6 ^b
LYS 13	10.8	10.2	10.6	10.7	10.6	0.2		
GLU 14	4.8	4.7	4.7	4.3	4.6	0.2		
ASP 20	6.4	3.6	6.8	3.4	5.1	1.8	<0	
LYS 21	11.9	10.1	11.0	10.6	10.9	0.8		
ASP 22	5.0	1.8	4.7	3.6	3.8	1.4		
ASP 24	4.5	3.0	4.8	6.0	4.6	1.2		
LYS 30	11.6	10.4	10.5	10.9	10.8	0.5		
GLU 31	8.8	5.3	9.9	6.9	7.7	2.1	5.0	6.5 ^b
ARG 37	12.2	12.2	13.4	12.7	12.6	0.6		
GLU 45	3.3	3.8	3.0	3.7	3.4	0.3		
GLU 47	4.1	3.9	4.4	4.4	4.2	0.2	4.1	4.6 ^b
ASP 50	4.4	4.1	4.5	4.4	4.3	0.2	4.7	5.2 ^b
GLU 54	4.5	4.6	4.3	3.7	4.3	0.4		
ASP 56	6.6	3.5	8.5	4.7	5.8	2.2	<0	
ASP 58	5.8	1.7	7.4	2.5	4.3	2.7		
ASP 64	3.3	3.1	3.8	5.7	4.0	1.2		
GLU 67	11.2	5.6	13.4	8.5	9.7	3.4	2.5	5.4 ^b
ARG 74	13.1	12.4	14.3	14.2	13.5	0.9		
LYS 75	11.3	10.5	11.3	11.4	11.1	0.4		
LYS 77	10.9	10.5	10.9	11.2	10.9	0.3		
ASP 78	3.7	3.8	3.4	3.3	3.6	0.3		
ASP 80	4.6	4.1	5.1	4.6	4.6	0.4	4.7	
GLU 82	2.5	3.8	3.9	3.9	3.5	0.7	3.4	
GLU 83	4.8	4.8	4.6	4.1	4.6	0.3		
GLU 84	5.8	4.9	5.4	5.0	5.3	0.4	5.2	
ARG 86	14.0	13.1	16.3	13.0	14.1	1.6		
GLU 87	4.6	4.6	4.6	4.2	4.5	0.2		
ARG 90	13.0	12.4	14.5	13.4	13.3	0.9		
ASP 93	7.3	2.4	4.5	5.6	5.0	2.1		
LYS 94	10.6	10.4	11.0	10.7	10.7	0.2		
ASP 95	5.2	3.8	5.3	2.7	4.2	1.2		
TYR 99	11.5	10.3	12.5	11.0	11.4	0.9		11.6 ^c
GLU 104	9.0	4.9	11.3	5.9	7.8	2.9		
ARG 106	12.9	12.5	13.8	13.3	13.1	0.6		
HIS107	6.1	6.0	5.8	4.7	5.7	0.6		6.2 ^c
GLU 114	4.7	4.7	4.8	4.0	4.6	0.4	4.9	
LYS 115	10.6	10.5	10.5	10.6	10.5	0.1		
ASP 118	4.2	3.5	5.2	3.9	4.2	0.7		

GLU 119	4.3	4.6	4.3	4.0	4.3	0.3	
GLU 120	4.2	4.1	5.0	4.5	4.4	0.4	4.7
ASP 122	6.0	4.9	5.6	6.7	5.8	0.7	4.8
GLU 123	4.9	4.6	4.9	4.0	4.6	0.4	
ARG 126	13.2	12.6	14.8	14.0	13.7	0.9	
GLU 127	5.2	4.8	5.2	4.3	4.9	0.4	5.1
ASP 131	4.4	4.5	5.0	2.6	4.1	1.1	
ASP 133	8.1	2.4	7.7	5.9	6.0	2.6	
TYR 138	17.7	11.6	13.4	12.9	13.9	2.6	
GLU 139	4.2	4.7	4.0	3.9	4.2	0.4	
GLU 140	10.5	3.3	10.7	6.5	7.8	3.5	

^a for the structure 1c1l (Ca²⁺ loaded CaM; homo sapiens) [114].

^b for calbindin, close homolog of CaM N-lobe, with structural alignment showing the following correspondence between CaM – calbindin residues as reported in (a) above: [Glu11=Glu5, Glu31=Glu27, Glu47=Glu48, Asp50=Glu51, Glu67=Glu65] [113].

^c from ref. 109. Therein, pK_a values of 3.4 and 4.3 were reported for aspartate and glutamate residues coordinating the ions, although exact assignment of residue indices was not possible.

Our findings indicate that at lower pH, E31 will be uncharged so that the side chain will not be stabilized by the Ca²⁺ ion, and therefore, will have a higher probability to occupy alternative conformations. PRS shows that such a local conformational change propagates to the linker region and beyond, to favor the compact forms. This finding is in agreement with the FRET experiments conducted at pH 5.0 versus 7.4 of Ca²⁺-CaM, where the distribution of distances between the fluorescently labeled donor (34)-acceptor (110) residues on either domain shifted significantly towards more compact conformations so that the extended conformation was almost entirely absent at reduced pH [48].

The pH effect has been noted as early as 1982, when the activation of MLCK by CaM was shown to occur in the pH range of 6.0–7.5 [117]. This is the range of pH where CaM is known to have the more rigid structure exposing the domain at the site on interaction. For the MLCK bound form, the catalytic activity exhibits a broad optimum from pH 6.5 to pH 9.0. This bound form is represented by the structure 2bbm in our set where PHEMTO calculations now find nine negatively charged residues whose pK_a values have upshifted to this range (D22, E31, D64, E67, D95, E104, D131, D133, and E140); all except D64 are EF-hand loop Ca²⁺ coordinating residues. Also, D122, which does not participate in EF-hand loops, but has a predicted pK_a in the physiological range (gray curve in Figure 6.(a)) is found by PRS to be important for the conformational change between 3cln → 1lin and 3cln → 1prw (Table 4) with non-

specific perturbation directions. It is plausible that this residue also acts as a local pH sensor for manipulating the conformations that favor closed form.

3.3. Molecular Dynamics Simulations

In this part, the data obtained from MD runs of the systems summarized in Table II will be given and analysis results will be discussed in detail in the subsections.

3.3.1. MD Simulation Trajectories of the 3cln and 1prw Structures (3CLN, 1PRW)

RMSD of the different parts of the protein are compared to the initial starting structures for each MD trajectory whose the details are listed in Table 2. In the 3CLN run, the RMSD of the protein increases to ~ 10 Å within 30 ns, and keeps this level until 200 ns of the trajectory. At the 200 ns, it further increases to ~ 12 Å and keeps this level for the rest of the simulation. Due to this finding, the MD trajectory of 3CLN run can be divided in three parts as the first 30 ns, the second part from 30 to 200 ns and the third part is the last 200 ns of the trajectory.

Initially, the constraints imposed on this highly charged protein by crystal packing are relieved in solution. However, neither the overall structure of the NTD nor that of the CTD show a change during the first half of the simulation as may be seen from the separate RMSD values of these regions (Figure 7, cyan and green). Only RMSD of the linker shows a slight increase from 0.7 to 2 Å within 25 ns of the trajectory. This is the result of the larger degree of conformational fluctuations of the linker compared to NTD and CTD as outlined in Chapter 1 [54, 90, 91, 93]. In the third part of the simulation, the conformational fluctuations of the NTD get larger which is accompanied by that of the linker. RMSD of the NTD shows an increase from 1 to 2.5 Å and linker RMSD increases from 2 to 3 Å. These levels are kept until the end of the simulation. However, CTD remain nearly the same configuration throughout the simulation as the RMSD is ~ 1 Å.

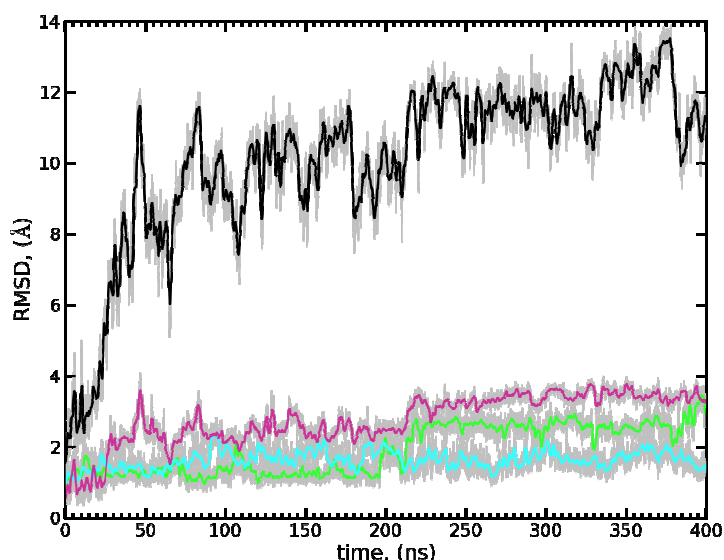


Figure 7. RMSD of the protein trajectory for 3CLN; moving average taken over 200 data points of the RMSD of the whole protein shown in black, NTD shown in green, CTD shown in cyan and linker shown in purple. The light gray dots show the actual RMSD of the full trajectory.

The difference in the conformational fluctuations from the three parts of the trajectory can be explained by the snapshots representing each part as shown in Figure 8.

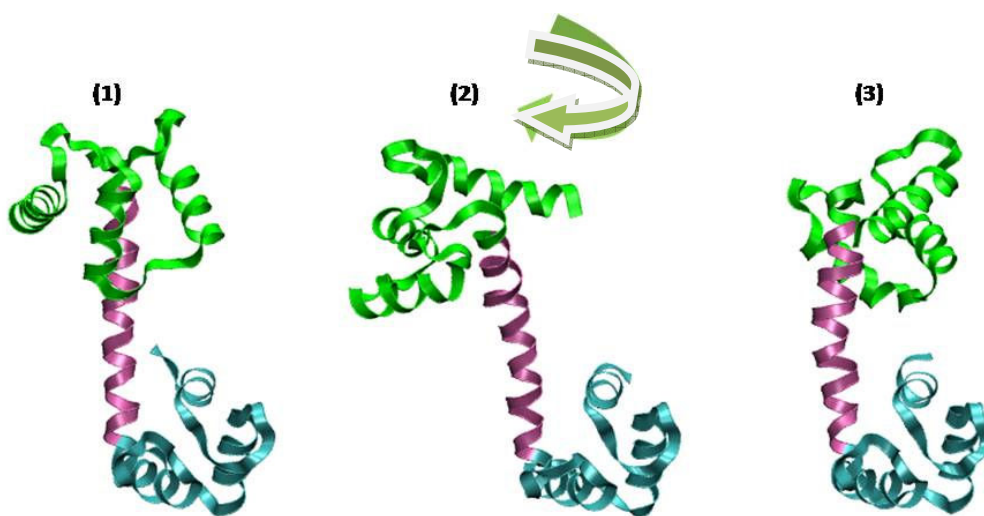


Figure 8. Representative snapshots taken from different parts of the 3CLN MD trajectory. The number representing each part is given at the top.

Each structure is aligned over the CTD as it has the least degree of conformational fluctuation and therefore can be assumed rigid during the entire simulation. RMSD change of overall protein in part 2 is due to the rotation of NTD with respect to the linker; rotation direction is shown by the green arrow in Figure 8. In part 3, NTD further rotates in the same direction and this change is observed in the overall

RMSD with a slight increase. Internal configuration in NTD changes due to the change in the orientation of helix I and helix II with respect to each other. These changes are also captured in the linker-end-to-end distance vs torsion angle plots whose features were explained in Section 2.7. In Figure 9, l_{linker} vs φ joint probability plot of the 3CLN simulation is shown. This plot is generated by the snapshots taken every 20 ps from the entire trajectory, making total of 20000 points.

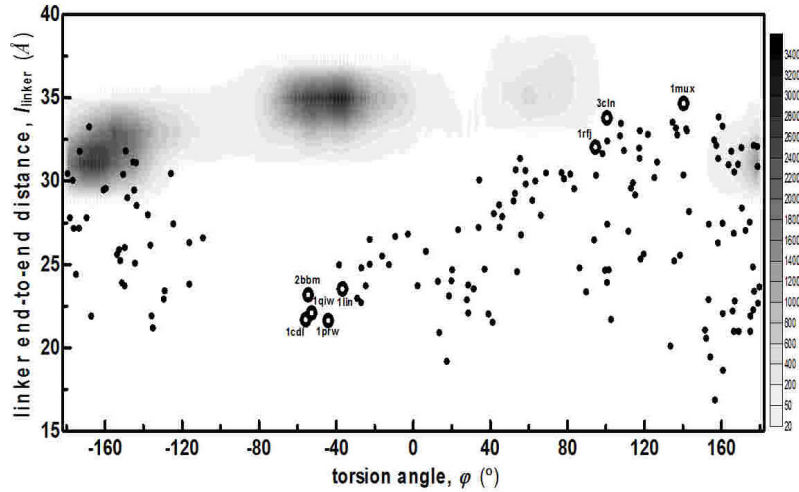


Figure 9. φ vs l_{linker} joint probability plot for 3CLN trajectory.

The (l_{linker}, φ) distribution shown in Figure 9 is calculated at 2° , 2 \AA intervals in φ and l_{linker} , respectively. The torsional angle of 3CLN structure spans large intervals of 76 ± 12 , -29 ± 19 , -130 ± 34 in parts 1, 2 and 3 of the trajectory, respectively. This finding is the main reason for the high RMSD fluctuations of the overall protein (Figure. 7). The end-to-end-linker distance resembles restricted behavior spanning narrow intervals of 33.3 ± 0.6 , 32 ± 0.8 and 30 ± 0.8 in parts 1, 2 and 3 of the trajectory, respectively. Thus, it is affected from the change in the torsional angle fluctuations, and shows a response by lowering its magnitude to compensate for the large change in the torsional angle. This hints at the presence of an energetic barrier in the end-to-end-linker distance or in other words there is an energetic barrier in the linker bending angle.

When we start the MD simulations from the compact structure 1prw, it is observed that the overall fluctuations are smaller on average compared to 3cln with a value around 6 \AA . This is due to its compact structure. Although this structure also moves away from its x-ray coordinates initially, NTD, CTD and linker act in a more coordinated fashion. The individual RMSD of the NTD; CTD, the linker as well as the overall structure, are shown in Figure 10.

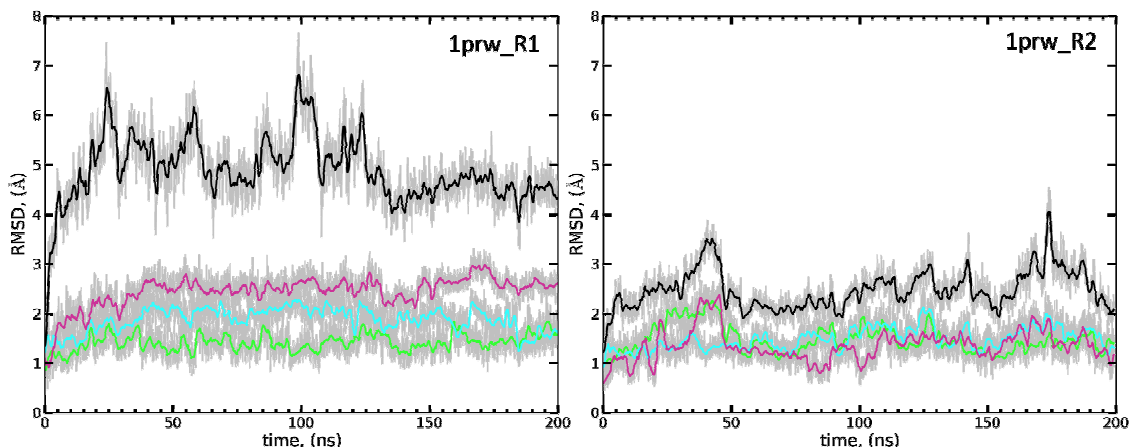


Figure 10. RMSD of the protein trajectory for 1prw (two independent MD trajectories, each 10000 snapshots); moving taken over 100 data points of the RMSD of the whole protein shown black, NTD shown green, CTD shown cyan and linker shown purple. The light gray dots show the actual RMSD of the trajectory.

The compactness of 1prw is maintained via the interactions between the interfaces of the two lobes in which Glu7 (N-lobe) and Glu11 (N-lobe) interact with Glu127 (C-lobe), Glu114 (C-lobe) and Glu120 (C-lobe) pairs with Glu14 (N-lobe). These acidic residues are found to be neutral at the pH of the x-ray experiment (pH=5.4, see Table 5 for pK_a values). Indeed, this finding suggests that these residues which are normally ionized at higher pH values interact unfavorably by forming hydrogen bonds [114]. When the crystal structure is placed in a low ionic strength, solution at 7.4 pH, as is the case in our work, the hydrogen bonds existing in the structure are not maintained fully as these residues are in their charged states in the solution. Due to the latter fact, new repulsive forces (unfavorable interactions) are formed across the two lobes and the crystal structure relaxes to a more open structure in the MD run. The regions sampled in the torsion and end-to-end linker distance space are shown in the l_{linker} vs φ joint probability plot in Figure 11.

The maintenance of the compact shape of 1pre is also supported by the fact that solvation fluctuations are not dominated by the highly negatively charged lobes. Two spheres of the same overall charge can be put into contact in water only if the surface charges are distributed discretely where 1prw interface charged residues lends support to this view [118].

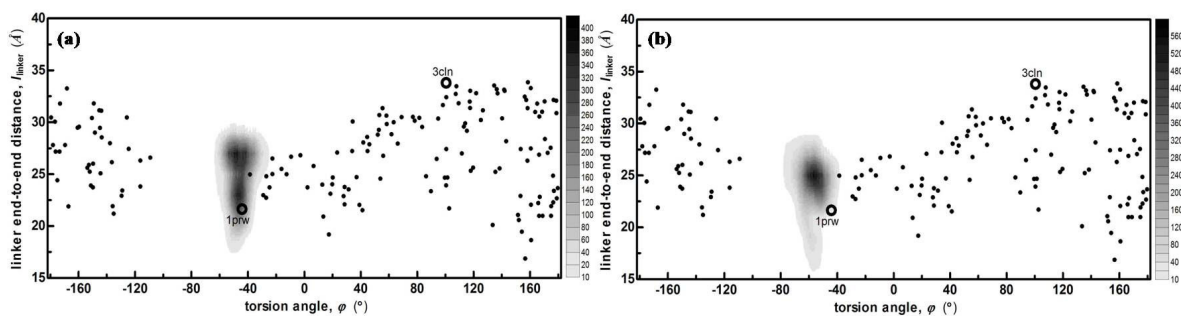


Figure 11. l_{linker} vs φ joint probability plot for 1prw runs. (a) 1PRW_r1 (b) 1PRW_r2.

The maintenance of the compact shape of 1pre is also supported by the fact that solvation fluctuations are not dominated by the highly negatively charged lobes. Two spheres of the same overall charge can be put into contact in water only if the surface charges are distributed discretely where 1prw interface charged residues lends support to this view [118].

Likely in 3CLN, the acidic and basic residues are kept in their default ionization states as they are in pH 7.4 and immersed in a low ionic strength solution. The two lobes with their highly charged surfaces present in the low ionic strength environment are the relatively far away from each other since charge repulsion between the lobes can not be screened. This leads to many conformations being sampled in the φ space. Thus, the protein, it cannot sample any conformation with l_{linker} value less than 30 Å due to the unfavorable energy of association of the lobes. The ionized states of the residues, possibly leading to additional repulsion, make the electrostatic component dominant and strongly oppose closer interdomain association [114].

3.3.2. MD Simulation Trajectories of 3CLN with 10 Residues Protonated (3CLN*)

PRS analysis on extended Ca^{2+} - CaM suggested that external perturbations on acidic residues with elevated pK_a values may cause the conformational change [17]. In fact, the system with 10 residues protonated (3CLN*) that are highly sensitive to lowering the pH leads to remarkable results. In 3CLN*, a different pattern in RMSD is observed. The overall conformation of the chain changes in two steps, first at ~ 15 ns, then at ~ 65 ns, while the rearrangements within the different units are relatively small (Figure 12). Thus, the main conformational change is due to the relative arrangement of the units with respect to each other. Similar to 3CLN, 3CLN* MD trajectory can also be divided in to 3 parts as part 1 (0-15 ns), part 2 (15 – 65 ns) and part 3 (65 -200 ns).

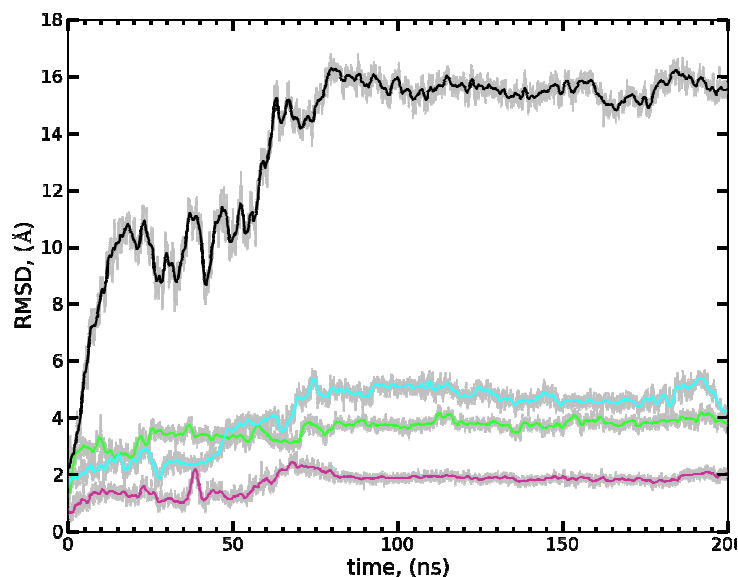


Figure 12. RMSD of the protein trajectory for 3CLN*; moving average taken over 100 data points. RMSD of the whole protein shown black, NTD shown green, CTD shown cyan and linker shown purple. The light lines show the actual RMSD of the trajectory.

Inspection of Figure 12 more closely, though, it is found that the CTD motions are relatively larger, while the linker maintains its original helical structure, but is observed to be slightly bent at some instants. This feature is significant and different when compared to 3CLN MD trajectory where CTD shows no remarkable change during the simulation. The linker being slightly bent at some instants is important because various authors have argued that the loss of the helicity in the central linker might be the source of conformational multiplicity [90, 93, 119]. The average RMSD of the linker is on the order of 3-4 Å in 3CLN and 3CLN*, respectively. In both cases, the initially extended linker does not sample bent conformations, despite the large degree of fluctuations accompanied by local unfolding events. The motions in 3CLN and 1PRW cases carry bending and torsional modes as discussed in reference 16.

Figure 13 shows the (l_{linker}, ϕ) distribution plot. In 3CLN* MD simulation, the extended starting structure samples the same region as 3CLN in the initial phases of the simulation, also providing a similar RMSD in the 10 Å range.

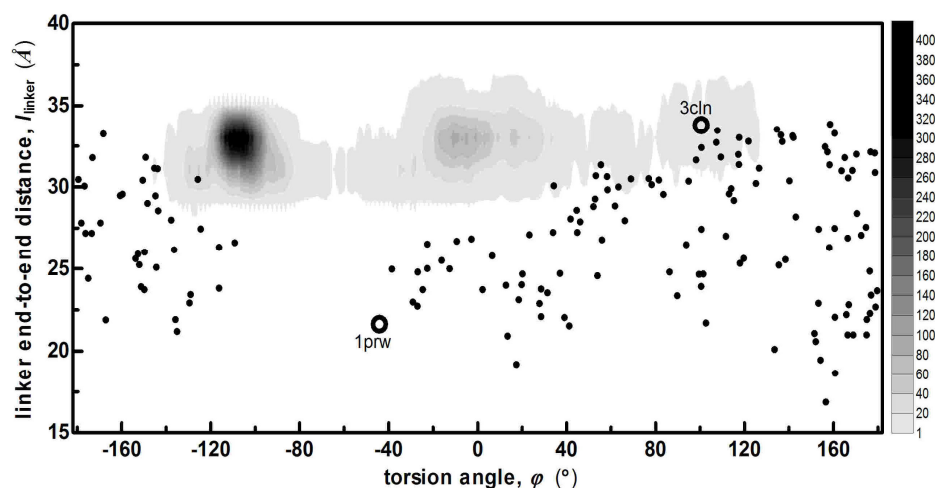


Figure 13. l_{linker} vs ϕ joint probability plot for 3CLN* run.

The torsion angle in 3CLN* MD run spans larger intervals than it does in the 3CLN run and in the range of $42 \pm 45.5^\circ$, $-10.5 \pm 26.6^\circ$, $-107.6 \pm 10.5^\circ$ for parts 1, 2 and 3, respectively. Fluctuations display larger deviations in parts 1 and 2 as the variations in the torsion angle is very high and the duration in these parts of the trajectory is short (15 ns in part 1 and 50 ns in part 2). In fact, in part 3 of the trajectory, it relaxes to a narrower region in the torsional angle when compared to that of 3CLN and stays there until the end of the trajectory. Therefore, the region that 3CLN* initial structure relaxes after 65 ns may be considered as a more stable point due to smaller torsional angle fluctuations. However, the end-to-end-linker distance shows no significant change where it stays in the intervals of $32.8 \pm 0.9 \text{ \AA}$, $32.6 \pm 1.1 \text{ \AA}$ and $32.2 \pm 0.7 \text{ \AA}$ in parts 1, 2 and 3 of the trajectory, respectively. This is an expected result, as the unfavorable attractions are altered by protonation of 10 residues distributed over the protein, the large variance in the energy of the system due to the torsional angle change is compensated. Hence, the alteration of the unfavorable attractions reflects the larger span in torsional angle which the deprotonated system lacks of. All of these observations show that the energy landscape is sufficiently adjusted by protonating the selected residues in 3CLN*. A sketch that represents the latter fact is shown in Figure 14.

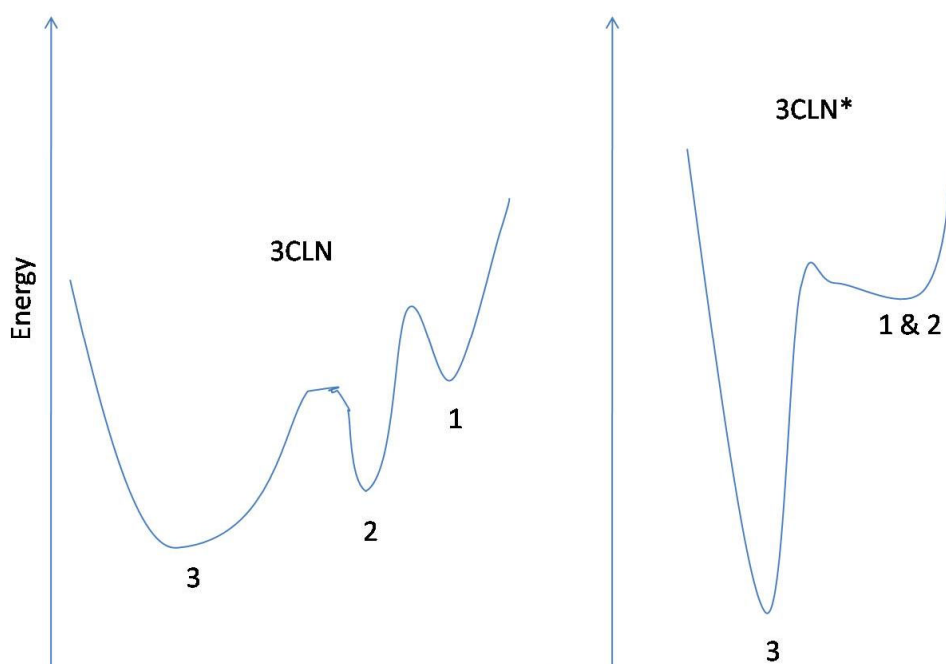


Figure 14. Schematic representation of the energy landscapes discussed according to the torsional angle change. The dotted lines show the energy level of the relaxed systems at the third part of the corresponding simulations.

3.3.3. MD Simulation Trajectories of 3CLN System; E31 Protonated (E31⁺)

PRS analysis between 3cln and other five target structures (recall Table 4) reveals E31 as an important. Hence, protonating a group of residues in 3CLN, that are mostly sensitive to pH changes, manipulates the initial extended Ca²⁺-CaM structure towards a new compact conformation that is not observed in the PRS target structures. In order to see the affect of the single residue protonation on the extended Ca²⁺-CaM, MD simulation of 3cln where only E31 protonated is performed. The details of the simulation are given in Table 2.

Two independent MD runs starting from the same initial structure were performed. The RMSD plots for the two cases are shown in Figure 15. Protonation of E31 that is residing in the NTD; one of the Ca²⁺ coordinating residues in EF hand I; introduces a reorganization in NTD which is reflected in RMSD of NTD. Due to this change, overall conformation of the protein change. However, the conformational fluctuation pattern is remarkably different than that of 3CLN*. Linker maintains its helical extended conformation throughout the simulation. Also, single protonation in NTD does not introduce any remarkable change in CTD as the conformational

fluctuations are on the average of below ~ 2 Å. From the RMSD plots, it can be concluded that single residue protonation does not result in any conformational change in the overall structure of the protein.

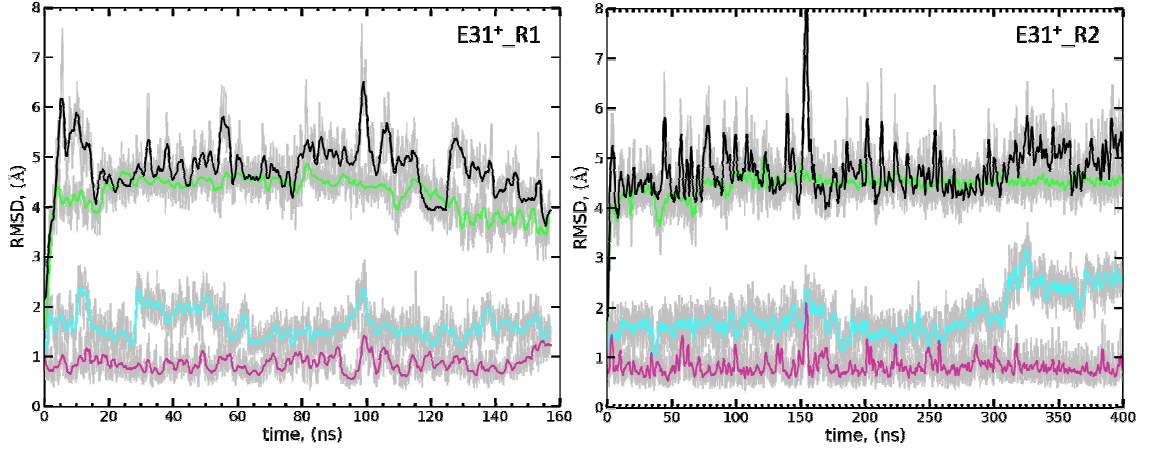


Figure 15. RMSD of the protein trajectory for $E31^+$ (two trajectories of 15000 and 20000 snapshots, respectively) ; moving average of the RMSD of the whole protein shown black, NTD shown green, CTD shown cyan and linker shown purple. The light gray lines show the actual RMSD of the trajectory. Moving average taken over 150 and 200 data points, respectively.

In Figure 16, the (l_{linker}, φ) distribution plot is shown. In $E31^+$ MD simulation, the extended starting structure samples the same region as 3CLN* samples in the very beginning phase of its trajectory. However, $E31^+$ does not escape to another region in the reduced degrees of freedom plot, while this region is one of the transients in the 3CLN* case. Torsion angle and end-to-end linker distance in $E31^+$ MD run sample a more constrained region. Thus, the protonation introduces restraint to the system so that the lobes stay distant from each other, keeping l_{linker} in the region 33 ± 0.7 Å while φ stays in the region $100.5 \pm 13.7^\circ$.

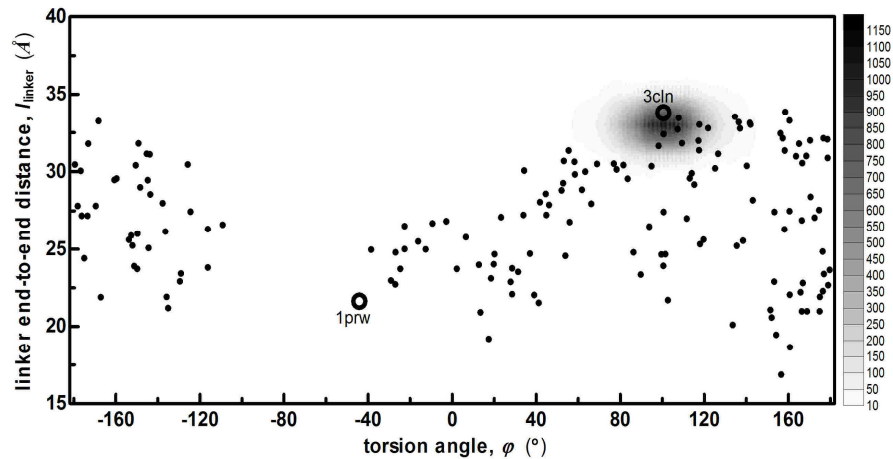


Figure 16. l_{linker} vs φ joint probability plot of $E31^+$ run.

In standard protonation, the region that the system is constrained to in torsion angle coordinate may be considered as an unstable region having a low energy barrier to the consecutive region sampled, since the initial starting structures in both 3CLN and 3CLN* MD simulations drift away in a fast manner. Hence, the single residue protonation stabilizes this region by lowering the free energy of this region and the barrier to the nearest state gets higher. Moreover, changing the charge of the residue by keeping the same topology prevents signals that are delivered to the CTD, as its large conformational fluctuations are the second important driving force leading to conformational change in 3CLN*.

Another important fact is that the interactions of some charge spacers in 3CLN* system are excluded as the result of the extensive protonation and this gives the system channels of relaxation to establish a “compact state” despite the linker remaining straight. The charged residues serve as electrostatic “spacers” in the structure. If one or more side chains of the residues is neutralized as they are at pH 5, their ability to act as electrostatic spacers would be curtailed leading to a more favorable energy to the interactions that are prohibited due to kinetic energy barrier. Actually, the 3CLN system has the tendency to relax into the more or less same region in l_{linker}, ϕ coordinates. The transition to the more compact form takes place within 220 ns for 3CLN and within 65 ns for 3CLN*. As a result in the wild 3CLN, the conformational dynamics are slower compared to the dynamics in 3CLN* system. Thus, the kinetic energy barrier between the new “compact state” and the initial extended state is lowered so the new state is achieved earlier in the 3CLN* and yet is stable.

3.3.4. MD Simulation Trajectories of 3CLN System; E31 Mutated (E31A)

In order to investigate the importance of residue E31, a new system is constructed by changing the topology of the residue with an Alanine mutation and three independent MD simulations are performed starting from the same initial system (details are given in Table II). The most drastic change in the extended conformation comes from one of the E31A MD simulations. The RMSD plots belonging to each run is shown in Figure 17. RMSD curves of NTD for each run show that a reorientation takes place by an increase of the RMSD value from 2 to 4 Å around ~ 25 ns of the corresponding trajectory. The most remarkable point is the change in the linker conformational fluctuations taking place at 60 ns of the E31A_R1 trajectory. In this run,

the RMSD value increases from 1.5 to 4 Å and stays there for about 100 ns. RMSD of CTD in E31A_R1 run stays on average at 1.7 Å, indicating no conformational change in this domain of the structure.

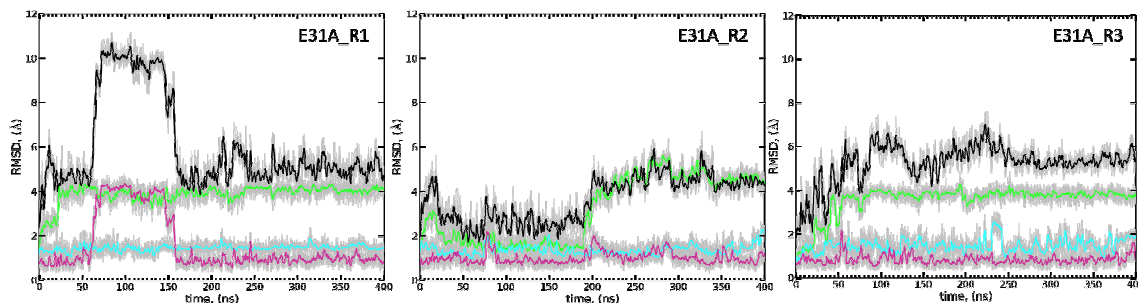


Figure 17. RMSD of the protein trajectory for E31A (three trajectories of each 20000 snapshots); moving average taken over 200 data points for each trajectory. RMSD of the whole protein shown black, NTD shown green, CTD shown cyan and linker shown purple. The light lines dots show the actual RMSD of the trajectory.

After 160 ns of the E31A_R1 run trajectory, the linker goes back to its earlier conformation while NTD stays in the new conformation. In fact, the conformational change in the NTD takes place in all three independent runs. In E31A_R2 trajectory with a delay of ~ 175 ns, in E31A_R3 trajectory more or less in the same time interval as it is in E31A_R1 run. However, this conformational rearrangement in NTD in these two latter runs does not result in the same large overall conformational change as in the first one. E31A_R1 run trajectory can be divided into two parts; part I consists of first 60 ns of the trajectory and trajectory interval between 160 – 400 ns, part II consists of the trajectory interval between 60 and 120 ns.

For these two runs CTD RMSD shows similar behavior by staying around ~ 2 Å compared to the CTD dynamics in E31A_R1 run. The details of the conformational change in the linker and NTD are appreciated when (l_{linker}, φ) distribution plots are inspected (Figure 18).

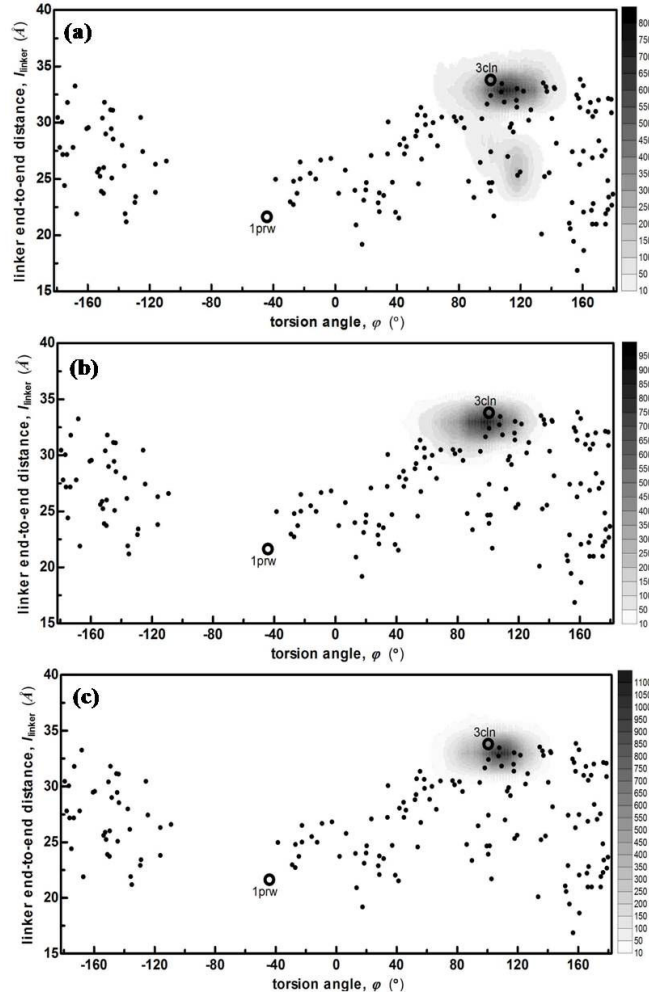


Figure 18. l_{linker} vs ϕ joint probability plots of E31A runs: (a) E31A_R1, (b) E31A_R2, (c) E31A_R3.

The initial structure drifts to the same region in the end-to-end-linker distance vs torsion angle coordinates as it does in the E31⁺ run. In the first interval of part I in E31A_R1 run, torsional angle of the structure is constrained around $100 \pm 14^\circ$ where the end-to-end-linker distance samples the values as if it is in the most extended state; with values of 33.0 ± 0.7 Å. In the second interval of part I, ϕ and l_{linker} spans the intervals of $113 \pm 14^\circ$ and 33.0 ± 0.6 Å, respectively. The conformational change that happens in between brings the system to a new state in the torsional angle coordinate, the average value of ϕ is increased by 13° . In part II of E31A_R1 run, structure begins sampling the same torsion angle region in the interval of $110 \pm 12^\circ$, while the end-to-end-linker distance establishes value 27.5 ± 2.3 Å. This abrupt decrease in linker distance results from the formation of a kink in the linker. Snapshots from the mentioned intervals are investigated, shown in Figure 19.

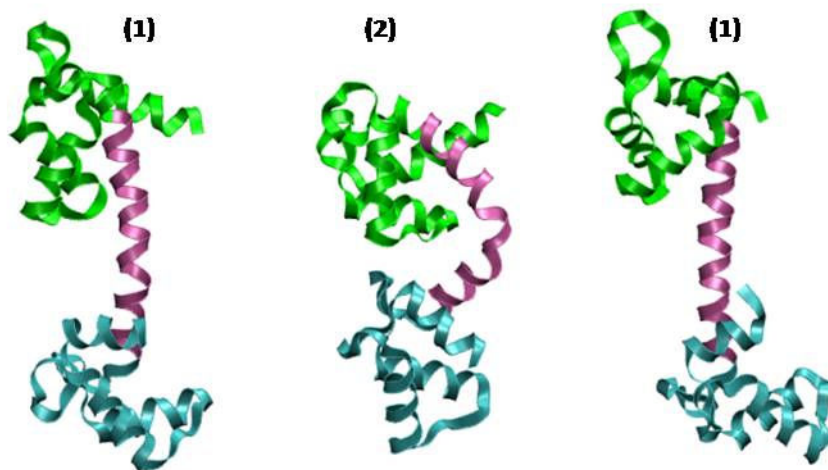


Figure 19. Representative snapshots taken from different parts of the E31A MD trajectory. The number representing each part is given at the top.

At 60 ns of the E31A_R1 trajectory, a salt bridge between residues E47 and E86 is formed and this is indicated with a decrease in l_{linker} from 33 to 27 Å. Following this event, at 80 ns the linker is bent from residue 81 and the l_{linker} values drops down to 25 Å. Around ~ 110 ns of the trajectory, the structure visits its most compact conformation with $l_{\text{linker}} = 25$ Å and $\varphi = 80^\circ$. Although the system samples a region in $\varphi = [60^\circ, 120^\circ]$ that is nearly the same as it is in the E31⁺ MD run of Section 3.3.3, the N- and C-lobe come in close proximity of each other implied by the sharp decrease in l_{linker} .

In order to make further inferences, time intervals of the transitions between the extended and compact (forward transition, t_f) and between compact and extended conformations (reverse transition, t_r) are examined in more detail. Snapshots are taken at every 100 ps in the trajectory points between 60-80 ns and 140-160 ns and shown on the $(l_{\text{linker}}, \varphi)$ coordinates in Figure 20. Note that the axes values have zoomed in to the relevant range.

The time interval of both transitions is very short as t_f and t_r takes place around 1 ns. The exact transition times are also deducted from this plot and they occur at 65.3 ns for t_f and at 146.2 ns for t_r . Both transitions use the same transient conformation to jump to the other state. Also, in both cases, the torsional angle shows an increase just before the transition. The transition between the extended and compact states is also examined via a tool called Geometrical Pathways [120, 121]. This tool utilizes geometric targeting (GT) method that has recently been introduced [120] as a rapid way to generate all-atom pathways from one protein structure to some known target structure. GT is based on the

philosophy that essential features of protein conformational changes can be captured by solely considering geometric relationships between atoms. The protein is modeled as a constrained geometric system, with constraints established to enforce various aspects of structure quality. These are: preserve covalent bond geometry, prevent overlap of atoms, avoid forbidden Ramachandran regions for backbone dihedral angles, avoid eclipsed side-chain torsional angles, and maintain hydrogen bonds and hydrophobic contacts [120].

We have generated 10 random pathways with a RMSD step size of 0.05 Å between the structures at 50 ns (extended E31A) and at 80 ns (closed E31A) from the E31A_R1 MD trajectory by using Geometrical Pathways in Biomolecules server [121]. The structures generated in the t_f pathway by Geometrical Pathways are plotted on (l_{linker}, φ) plane with the median of the pathway and the standard error bar along both axes. In Figure 21, it can be seen that the pathways produced covers the pathway that is sampled by MD. In GT generated pathways, energetics is neglected and the domains in a GT pathway move simultaneously at each step as there is no geometric reason for a sequential movement. We note that predicting relative timing of events is not predicted by GT.

The similarity between the pathways may also arise from the highly constrained nature of proteins in general. Geometric considerations such as covalent bonds, non-overlapping atoms, maintenance of hydrogen bonds, etc. do in reality severely limit accessible conformational space and restrict the possible ways that a protein can move from the initial to target state. Another reason for the pathways being similar can be the sequential movement of the two domains in the MD trajectory.

The transition elucidated from MD is compared with the random pathways that are generated by using Geometrical Pathways. These random pathways are widely distributed along the interdomain torsional angle and all have narrow distribution in end-to-end-linker distance axis.

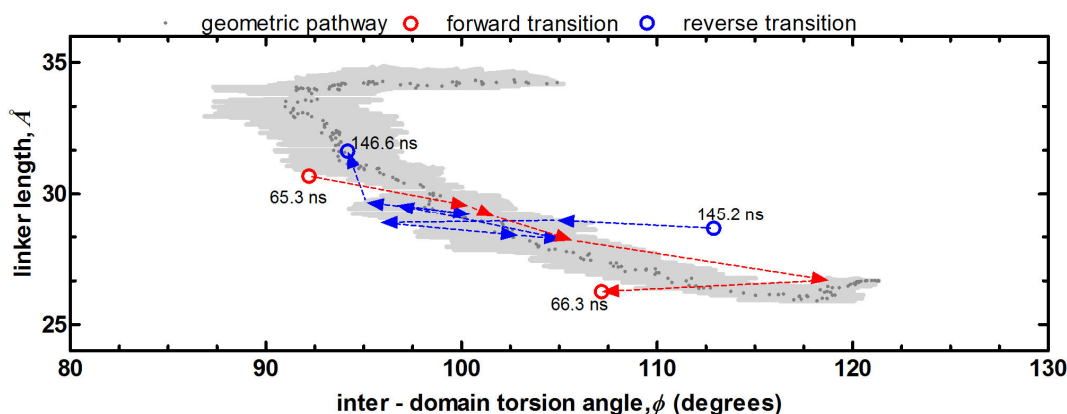


Figure 20. The dark gray points are the median of the pathways generated by FRODA and the light gray shaded area is the standard error bars along each axis. The initial (50 ns) and target (80 ns) structures that are fed to Geometrical Pathways are shown with black outlined empty circles. Forward transition (64.3 – 66.7 ns) and reverse transition (145.2 – 147.2 ns) paths.

Since the transition only takes place in one of the runs out of three in the simulation set, this event can be considered as a rare event. In order to understand the main difference between these independent E31A runs, in the following paragraphs, focus will be given to more dynamic features that can be deduced from MD simulation data whereby the angles between the helices in each EF hand motif were monitored throughout the trajectory.

We first investigate relative rearrangement of helices within each EF hand motif, In Figure 21, the angle trajectory between selected helices within each EF hand is shown. Helices I and II in EF hand I, and helices III and IV in EF hand II are both initially posed at about right angles to each other, values between 80-90°. In run 1, they both show a sharp decreasing trend, the former until 50 ns and the latter until 30 ns. Also, EF hand IV makes a closing motion until 60 ns of the trajectory and this may be interpreted as an evidence for NTD and CTD communication. This fast event is not observed either in E31A_R2 nor E31A_R3. However, in E31A_R2, while EF hand I makes a closing motion at 240 ns, EF hand II does not respond with a closing motion. In E31A_R3, EF hand I shows this type motion after 90 ns and again EF hand II does not respond. Another fact that is remarkable in E31A_R1 trajectory is that EF hand I makes an opening motion at 140 ns and reaches to its second maximum value through its trajectory. This time exactly corresponds to the time interval of the reverse transition. We further investigate how consecutive EF hand motifs within each domain interact (Figure 22). The angle between helices I and III gets larger meaning that EF I and EF II

make an opening hinge motion relative to each other. EF hand II thus gets in proximity of the linker. Therefore, residue 47 which resides on helix III on EF hand II has the propensity to make the salt bridge with residue 86.

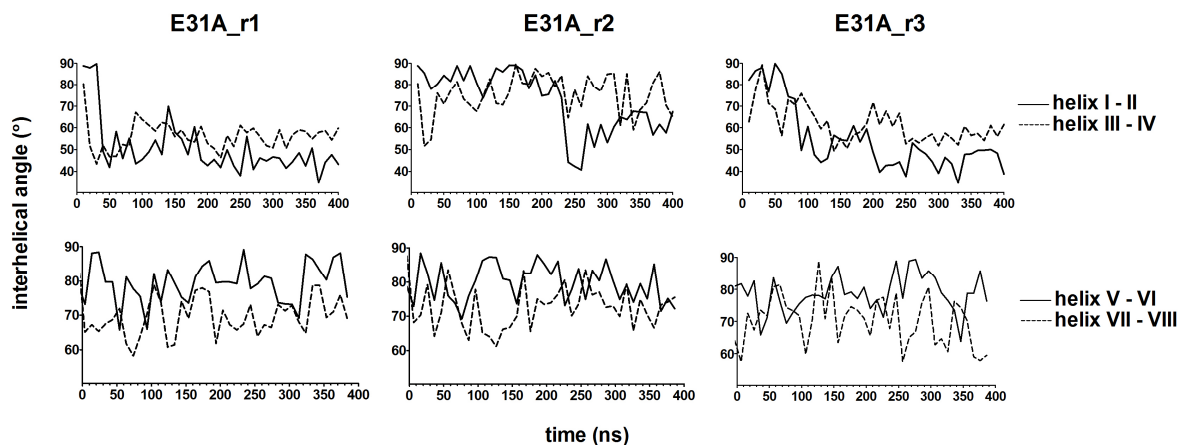


Figure 21. Inter helix angle trajectories. In the upper panel, angle trajectories for the NTD in each run are shown. The bottom panel shows angle trajectories in CTD.

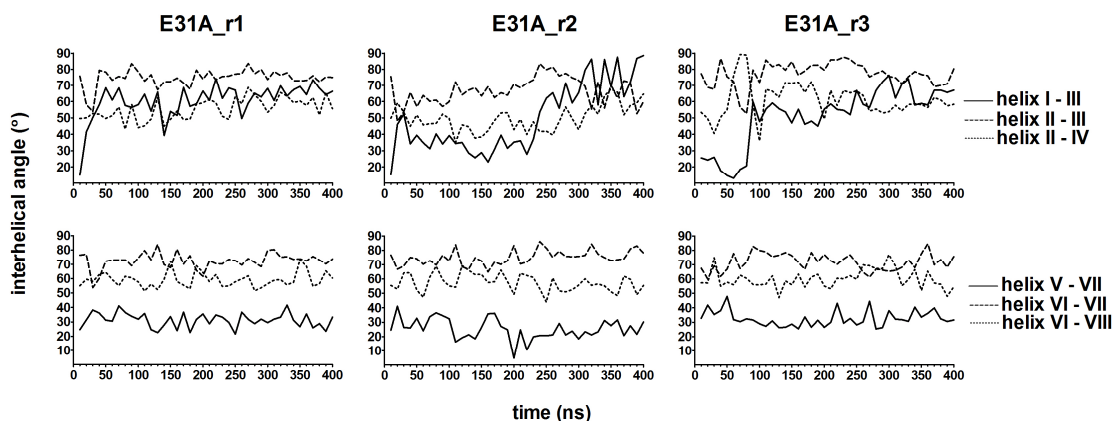


Figure 22. The angles between the helices in consecutive EF Hands throughout each MD run. Top panel shows the ones for NTD and bottom panel shows the ones for CTD.

Finally, when these trajectories are compared to the ones for the 3CLN MD simulation, it is observed that due to the lack of E31A mutation in the helix II, the angles between helices I and III and helices II and IV fluctuate less in the original system. In CTD, the fluctuations of the angle between helices V and VII are suppressed (Table 6).

Table 6. Average inter-helical angles of helices V-VI and standard deviations.

Run Label	Angle
3CLN	38 ± 11
E31A_R1	31 ± 5
E31A_R2	25 ± 7
E31A_R3	33 ± 5

3.3.5. MD Simulation Trajectories of 3CLN at 150 mM IS

Experimental and computational studies show that at low IS, Ca^{2+} -CaM remains in the extended conformation [122, 123]. Also, salt screening has been known to alter polar interactions, thus making Ca^{2+} -CaM available for target molecule binding. In order to provide additional clues into how environmental conditions effect conformational change, MD simulation at 150 mM was conducted (Table 2). Its RMSD plot is displayed in Figure 23.

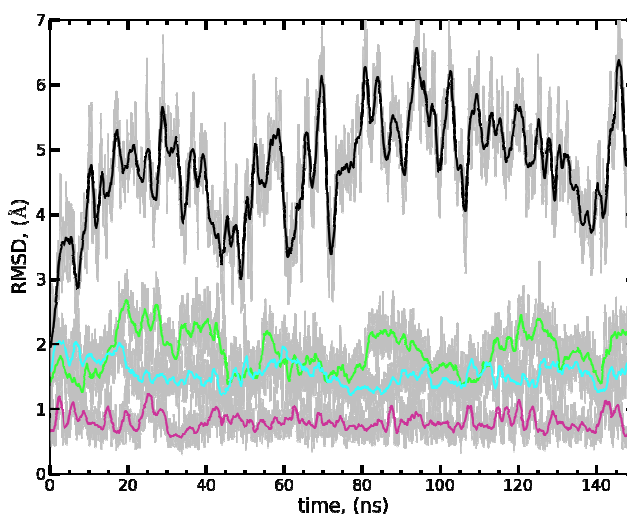


Figure 23. RMSD of the protein trajectory for 3CLN^{150mM} (trajectory of 7500 snapshots) ; moving average taken over 75 data points. RMSD of the whole protein shown black, NTD shown green, CTD shown cyan and linker shown purple. The light gray lines show the actual RMSD along the trajectory.

At the 10 ns point of the trajectory, the RMSD value for the whole protein increases to 4.5 Å and fluctuates around this value until the end of the simulation which was not the case for low IS 3CLN run (section 3.3.1). This indicates that the protein fluctuates around more or less the same conformations and there is no conformational

change within this time window. This is also confirmed with the RMSD fluctuations of the linker that fluctuate around 1 Å. The linker stays in the extended state during the whole MD simulation which is another major difference from the low IS 3CLN run. Moreover, the RMSD fluctuations of the NTD are more than that of the CTD. Increasing salt concentration results in more local rearrangements in the NTD. This finding is consistent with other modeling and experimental measurements that indicate a greater inherent plasticity within the NTD, a property that has been suggested to be important for target protein recognition. The same behavior for NTD is seen in the low IS 3CLN run after the 200 ns point of the trajectory, where the conformations relax into the region of NMR detected ones (Figure 9).

From the RMSD plot, it is apparent that introducing more salt to the environment does not lead to any conformational change in the protein. The (l_{linker}, φ) distribution plots demonstrate more information on the effect of the salt concentration on the protein conformation (Figure 24).

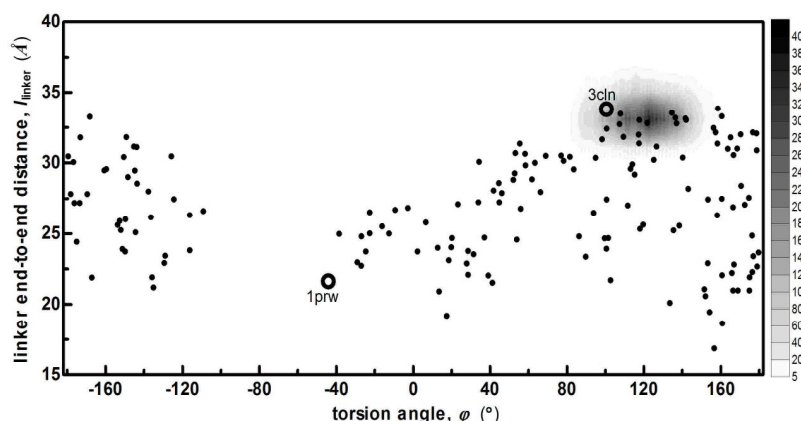


Figure 24. l_{linker} vs φ joint probability plot of 3CLN^{150mM} run.

Typically, salt screening introduces a similar effect to those in E31A and E31⁺ runs and constrains φ and l_{linker} to a narrower region. The torsional angle spans values of $120.5 \pm 14.5^\circ$ and end-to-end linker distance spans 33.3 ± 0.7 Å. Torsional angle is constrained to a region which has drifted from the torsional angle interval of the E31A and E31⁺ runs by $+10^\circ$ whereas, l_{linker} is in the most extended conformation.

Large fluctuations in the angle trajectories of the EF hand helices in NTD are consistent with the RMSD behavior of NTD (Figure 25). The angle between helices II and III, which is an indicator of the EF hand I and EF hand II dynamics, takes values

below 60° , in 45% of the trajectory. This means that the EF hands are positioned closer to each other, which was not observed in any of the other MD runs. Moreover, EF hand I stays in a more open conformation while EF hand II gets into a more closed conformation when compared to all other MD runs. EF hand III fluctuates less in comparison to other runs while EF hand IV goes to a more open conformation as if in E31A and E31⁺ runs and not like in 3CLN run in low IS. Since these two features are different from the other runs, this information reveals that salt screening introduces different alterations in the NTD even though reduced conformational space is constrained to similar regions.

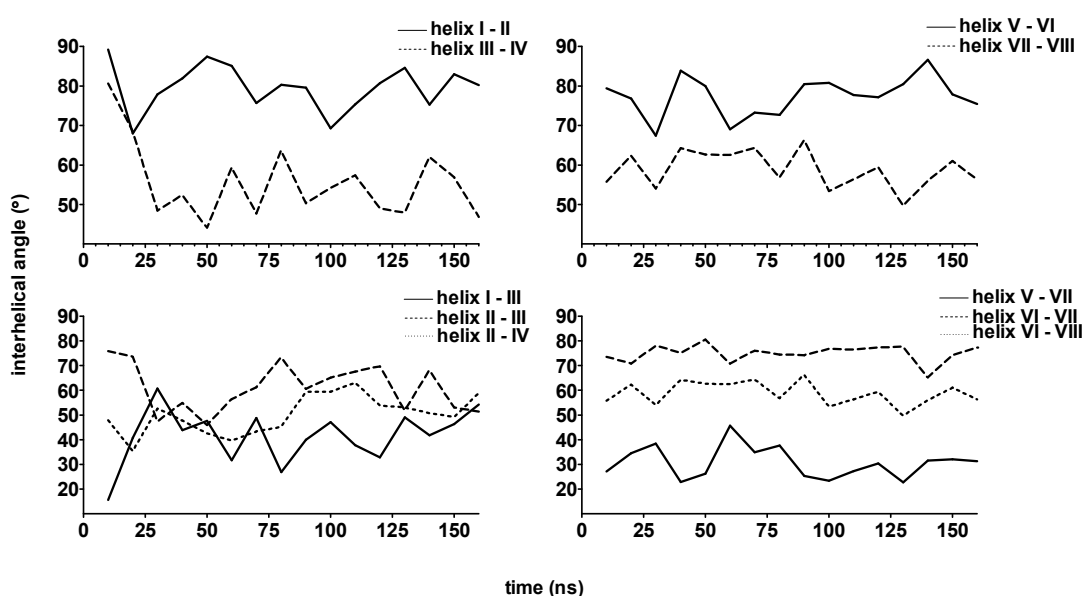


Figure 25. Angles between the helices in the EF Hands (top panel) and between the consecutive EF hands (bottom panel) throughout the MD trajectory of 3CLN at 150 mM IS.

3.4. Response of CaM to Local Displacements

The MM scanning methodology used in this work relies on introducing a perturbation at a selected residue in the form of a displacement, and extracting the response of the protein after energy minimization. We have carried out the MM scanning with moves of sizes 0.25, 0.5, 0.7, 0.86, and 1.0 Å, and we have calculated the overlaps from the reproduced response vectors to the target structure 1LIN. In a previous MM study conducted with chymotrypsin inhibitor 2, the size of the displacements was increased to 2.5 Å and a different perturbation-response pattern was observed [35]. It was concluded that for perturbations up to about one bond length, the system remains in the limits of linear response regime.

In the same study, the effect of the direction of the perturbation to the response pattern was also investigated. The perturbation-response maps from three different scans: (1) a displacement of 0.5 Å in each of the x -, y -, and z -coordinates is introduced, resulting in an overall displacement of 0.87 Å; (2) same as (1), but with displacements of -0.5 Å; and (3) uniformly distributed randomly selected moves in each of the x -, y -, and z -coordinates that result in a displacement of 0.87 Å were compared. Same pattern of perturbation-response was captured in all cases [35]. This finding implies that moves are readily localized, and conducted to the rest of the molecule through the similar mechanisms. This localization is coordinated by the packing geometry and packing density so that the system responds to an external perturbation with its topology. That the system responds with its intrinsic properties is another indication of linear response.

We choose moves of size 0.5 Å in all the MM calculations due to higher overlap values compared to the other move sizes. The MM calculations are randomly repeated for 25 times. The random directionality of the perturbation is expected to be consequential on the response pattern of the structure. Therefore, first the effect of the directionality of the displacement is investigated. The perturbation-response maps from the $\Delta\mathbf{L}$ matrix of equation 16 are reproduced for the 25 trials. In Figure 26, three samples of these matrices are shown. The pattern of the perturbation-response maps is similar in all runs.

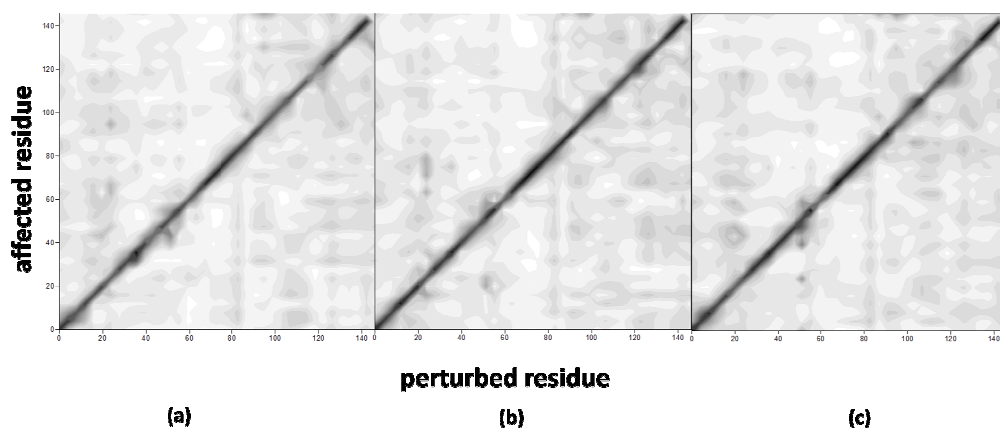


Figure 26. Molecular Mechanics perturbation-response maps of 3CLN for different directions (a, b, c) are shown.

Then, overlaps for each pair in the study set is calculated for these 25 MM runs. Table 7 shows the results from MM analysis. Local perturbations introduced by MM protocol reveals that a displacement perturbation on residues 117 and 17 may initiate the

conformational change towards proteins that are grouped in the lower part of the l_{linker}, φ plot in Figure 5 (1prw, 1lin, 2bbm, 1cdl, 1qiwi). These two residues are Threonine and Serine, respectively.

Table 7. Best overlap values obtained for proteins studied by MM .

Protein pair	Best MM Overlap / Residue ^a	Residue ^b
3cln/1lin	0.64 / 117	110, 115, 117
3cln/1prw	0.64 / 117	17, 115, 117
3cln/1qiwi	0.62 / 117	17, 110, 115, 117, 118
3cln/2bbm	0.56 / 117	17, 18, 115, 117
3cln/1cdl	0.58 / 117	17, 110, 115, 117, 118
3cln/1rfj	0.67 / 42	34, 37, 38, 40, 42, 44
3cln/1mux	0.57 / 88	84, 88

^a best results from 25 independent MM runs initiated from 3cln structure at 20 ns.

^b residues appearing in all of the runs that have overlaps within 0.05 of the maximum overlap value.

For 1rfj, residue 42 and for 1mux residue 88 reveals best overlap for the initiation of the conformational change. When we look at the residues that emerge in all of the runs that are shown in Table 7, for the protein group 1lin, 1prw, 2bbm, 1cdl we see that the residues appearing are mostly the same. Those on the NTD reside on helix I and the ones that are on the CTD reside mostly on helix V and VI. Therefore, the perturbation that results in conformational change for this group of residues can be introduced either from the NTD or CTD. Conversely, the local perturbation should be given to the residues in NTD for steering the conformation towards 1rfj or 1mux.

From the perturbation response maps the most affected residues can be calculated and can be ranked accordingly. This reveals supplementary information for the nature of the response. For the 25 independent MM runs, the most affected units are Ca^{2+} in EF hand I, followed by 27, the third most affected are residues 63 and 64 and the fourth most affected residues are 63, 67 and 74. With each rank, new residues are

included to the results and the data scans more regions in the protein. Therefore, the distance analysis does reveal important information on the nature of the response supplementary to the directionality measured by overlap. Whatever the direction of the displacement given to the residues, Ca^{2+} in EF hand I responds the most to the perturbation, then residue 27 that is in proximity to the ion then the response signal is mediated to residues 63 and 64 then in the fourth rank residue 74 is affected which resides in the beginning part of the linker.

3.5. Constant Force Perturbation Introduced by Steered MD

In this work, SMD was used to investigate the effect of the force direction that is determined by PRS on the extended Ca^{2+} -CaM; pdb code 3cln. PRS revealed that residue 31 has the tendency to initiate the conformational change between 3cln and other compact forms under the linear response regime assumption. We test this assumption by steering E31 C_α atom in the force direction that gives the best overlap with the target structures. Therefore, force is applied to the 3cln structure whose the coordinates are taken from the 20 ns point (3CLN^{20ns}) of the 3CLN MD trajectory. C_α atoms of residues 24 and 57 are kept fixed in space to avoid translation of the protein along the given force direction during the SMD simulation. After making test runs with force values between 1 to 250 pN, 80 pN is selected as the force magnitude in all SMD runs.

In Figure 27, the force (in (b) red arrow) that leads to maximum overlap for 1lin and the response (in (b) green arrows) given to this perturbation are presented. Also in the same figure (in part c) the linear path between the 3cln^{20ns} (SMD starting structure) and 1lin is presented with the SMD path that is obtained by steering E31 C_α in the shown force direction. Initially, 3cln^{20ns} structure is steered along the linear path by decreasing the torsional angle shown by the first 3 points on the SMD path. Then the structure cannot compensate the payoff for this sharp decrease in torsional angle and jumps back to higher torsional angle state and fluctuates around this point and the initial starting point in the $(l_{\text{linker}}, \varphi)$ plane.

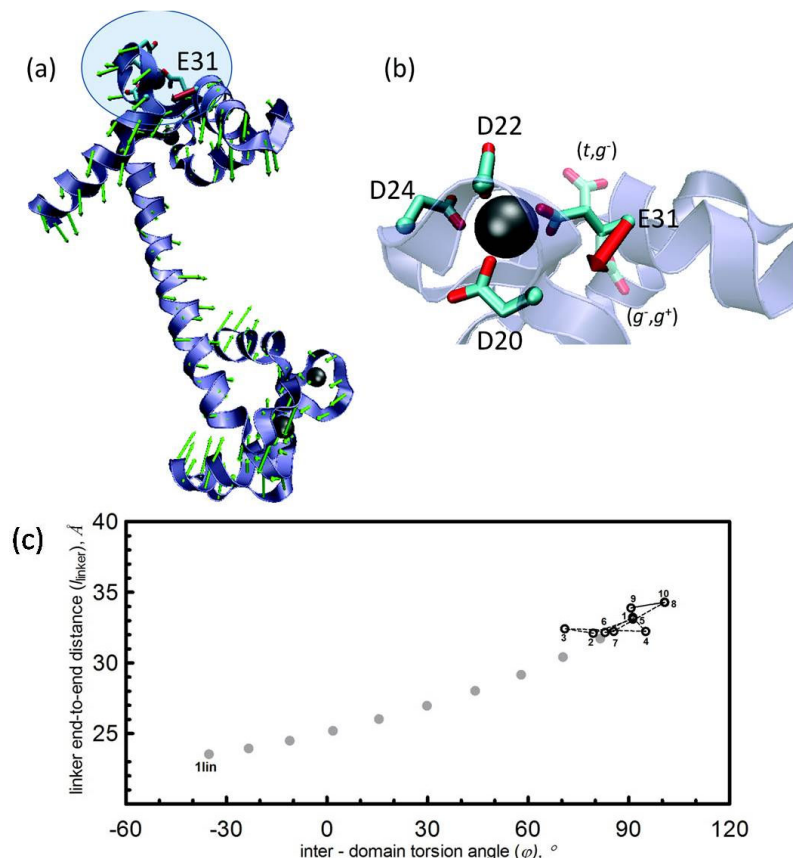


Figure 27. (a) Best PRS prediction of the displacement vectors (green) belonging to the 3cln to 1lin conformational change, overlaid on the initial structure. The main motion is a bending of the two lobes accompanied by rotation of the linker, where motions are especially accentuated in the region of residues 75–90. (b) Coordination of the E31 related Ca²⁺ ion. The side-chain (χ_1, χ_2) angles of E31 are in (t,t) conformation. Two alternate conformations of E31 which do not clash with any other heavy atoms in this conformation are also shown as transparent traces. These have (t,g⁻) and (g⁻,g⁺) conformations for the (χ_1, χ_2) angle pair. The thick red arrow represents the best perturbation direction of E31 in both figures. (c) Linear path between 3cln^{20ns} and 1lin shown with gray points, the path followed by the 3cln^{20ns} during the SMD is shown black points; each point is taken at 200 ps from the 2 ns long SMD trajectory.

Then it starts to sample a region which is not on the linear path. This implies that there is an energy barrier which can not be overcome by giving a mechanical perturbation to the 3CLN system. This may be due to energetic payoff of the mechanical force which is lower than that of the non bonded interactions in the protein. Thus, the applied force does not steer the starting conformation to a conformation that is not accomplished with the MD run (recall Figure 9).

It is known that a large conformational change occurs when residue 31 is mutated to Alanine (in one of the three runs of E31A system). The system accomplished a compact conformation that resides on a different part in the (l_{linker}, ϕ) plane (recall

Figure 18). SMD is applied to the structure at 50 ns from the E31A_r1 (E31A_r1^{50ns}) trajectory. The direction of the steering force is received by performing PRS between the structures at 50 ns and 80 ns from E31A_r1 trajectory. The best overlap giving residue is not residue 31 but residue 105. Residue 27 and 54 C_α atoms are kept fixed in space during SMD simulation. The change in the residue is expected since the topology of the residue along with the cross-correlation information of the system has changed. Also, after knocking out the transition initiating residue in NTD, a residue in CTD which resides at the topologically equivalent position to E31 is revealed to initiate the transition. 100 runs of SMD are performed by steering residue 105 in the selected direction and 20 of them resulted in a compact structure with a kink in the linker at residue 81. The RMSD between the compact structure that is received at 80 ns of the E31A_r1 trajectory and a compact structure that is accomplished in one of the SMD runs is shown in Figure 28. RMSD between these structures is 5 Å.

After investigating the SMD runs for both 3cln and E31A_r1^{50ns} structures, it can be speculated that mutating residue 31 to Alanine to a different topology leads to a lower energy barrier between the extended and compact states so that the energy barrier can be overcome in SMD. Another scenario can be, the energetic barrier between the extended and compact states stays the same while the mutation brings the system to a point in free energy landscape which is closer to the barrier of transition so that in some of the SMD runs as well as in one of the E31A MD runs, the initial system accomplishes the transition to a compact state.

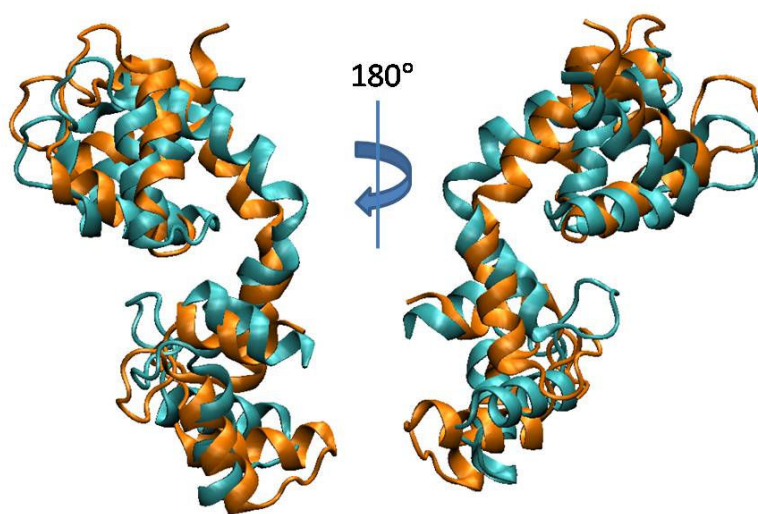


Figure 28. Snapshots of the compact structures accomplished by E31A_r1 (80ns, cyan) and SMD (2ns, orange).

Torsional angle and end-to-end linker distance data are calculated for all of the SMD runs starting from E31A_r1^{50ns} structure. It is observed that in each case the initial structure is not steered to the same end conformation, where it resides in different points is (l_{linker}, φ) plot (data not shown). This result also supports both energy barrier scenarios aforementioned. Thus, the energy landscape is altered by mutating E31 so that mechanical perturbation results in transition to compact structure in 20% of the sample runs which was not observed in 3CLN SMD runs. The occurrence of the transition is relatively rare since not every SMD run results in the conformation with the bent linker. In this work, SMD protocol is used in other way than that was utilized in literature and revealed complementary information to PRS.

CHAPTER 4

Discussion

4.1. Features of the Conformational Change Revealed by PRS

We may group the target protein structures according to the findings that resulted from PRS runs using 3cln^{crystal} structure: Group I consists of 1lin, 1prw, 1qiw, 2bbm and 1cdl, where the change is best captured by perturbation residue E31 and its immediate neighbors. Group 2 has 1rfj and Group 3 has 1mux. From the RMSD values in Table 3, we also know that the internal atomic rearrangements in the C-lobe is almost non-existent and those within the N-lobe are in the range of 0.4-3.3 Å. In contrast, the RMSD between the structures may be as high as 16 Å, which must mainly be coordinated by the flexible linker. This viewpoint is supported by NMR results whereby multiple conformations of Ca²⁺-CaM were discussed from the perspective of the linker [57]. That model yields compatible solutions to the experimentally measured nuclear coordinate shifts and residual dipolar couplings if the linker is modeled flexibly in the range of residues 75-81, while N- and C- terminal domains are assumed to be rigid. The analysis also suggests that all sterically non-hindered relative conformations of the two domains are not equally probable, and that certain conformations are preferred over others in solution.

PRS captures the conformations which are located far apart in the coarse grained conformational space (Figure 2) by giving perturbations to the same single residue. These perturbations are direction specific. In Figure 27 (a), we present the perturbation (red thick arrow) and the response (green arrows) that leads to the maximum overlap for 1lin. It is observed that the response is a simultaneous bending of the two lobes towards each other accompanied by the twisting of the linker. Also, the diagnosis of the PRS method is that the main motion governing conformational change is a collection of the slow and moderate modes [17], and that they may be best described as a twist and a bend for Group I molecules. It also demonstrates that it is possible to simultaneously

induce them via single residue perturbation with the correct directionality. These observations are supported by a Monte Carlo study on helix models that suggest applied torques along with constraints on the ends of α helical regions lead to a nonlinear coupling between the bending and extensional compliances [95].

4.2. E31 is a Signaling Residue for Global Communication in CaM

E31 consistently emerges as an important residue in manipulation of 3cln [96] implicates that E31 occupies an important location affecting the dynamics in apo CaM as well as partially or fully Ca^{2+} loaded CaM. E31 was implied to be involved in interdomain interactions of Ca^{2+} -CaM in an EndoGluC footprinting study [42]. EndoGluC proteolysis specifically cleaves at non-repeating glutamate sites of which there are 16 in CaM. The results point to E31 as a unique site involved in cooperative binding between the two domains. Cleavage at this site does not occur in apo and fully loaded states, but is significant in the partially loaded state. This induces susceptibility of E31 to cleavage is remarkably correlated to the induced protection from cleavage at E87, implicating that the observed changes are not local and possibly cooperative.

Furthermore, a structural homolog of the N-terminal domain of CaM is represented by troponin C (TnC). The structural alignment between TnC (PDB code 1avs: residues 15-87 and CaM (PDB code 3cln: residues 5-77) were performed and it yields a RMSD of 1.0 Å. Seventy percent of the aligned residues are identical and 88% are homologous, making TnC a viable model for the NTD of CaM. Ca^{2+} loaded structure of TnC has been determined at 1.75 Å resolution [124]. Moreover, single site E41A mutation in this protein and analysis by NMR indicates that there is direct coupling between binding of calcium to this particular EF-hand motif and the structural change induced [39]. E41 was found to be strikingly unique in control of TnC motions which is shown to single handedly lock the large conformational change whereby several residues have to move by more than 15 Å. The structural alignment of CaM and TnC reveals that not only E41 of TnC and E31 of CaM occupy analogous positions in terms of Ca^{2+} ion coordination, they also both have the same overall EF-hand motif structure. Therefore, it can be assumed that the critical role attributed to E41 in TnC is transferable to E31 in CaM.

The similarity of these two residues is also corroborated by E31K mutations which do not lead to apparent binding affinity changes of Ca^{2+} to CaM [39], as also

occurs in the E31A mutation of TnC [39]. Conversely, E \rightarrow K mutations in the other three equivalent EF hand motif positions of CaM (E67K, E104K, and E140K) lead to the loss of Ca²⁺ binding at one site [41]. Furthermore, E31K mutation has wild type activation on four different enzymes; smooth and skeletal muscle myosin light kinase (MLCK), adenylylcyclase, and plasma membrane Ca²⁺-ATPase, while other mutants in the equivalent positions have poor activation[40]. Double mutants of these sites suggest a tight connection between loop I and loop IV, and this coupling is possibly mediated by the linker [125], since there is no Nuclear Overhauser Effect (NOE) detected between NTD and CTD residues [126].

The connection between E31 location and the linker was later shown by a comparative MD study on Ca²⁺ loaded CaM versus CaM where the Ca²⁺ ion in EF-hand loop I is stripped from the structure. This study reveals that although the former is stable in elongated form during the entire course of the simulation (12.7 ns), the lack of this particular Ca²⁺ ion leads to structural collapse of the two domains at 7.5 ns [96]. This change was observed to follow the loss of helicity in the linker region.

In Figure 27, in order to further investigate the connection between loop I local structural changes and the linker, the response profile of the perturbation that leads to the largest overlap between the experimental and predicted displacement profiles, $O^{31} = 0.72$ is displayed. The overall bending of the two lobes towards each other is clear. Also, the response is small in the first 1/3 portion of the linker, while it is magnified in the bottom 2/3 portion, past R74 around which the linker has been noted to unwind even in early and much shorter (3 ns) simulations of CaM, possibly facilitating the reorientation of the two calcium binding domains [91]. In fact, more recent MD simulations of length 11.5 ns at physiological IS revealed that the central helical region unwinds at ~ 3.5 ns, while keeping the radius of gyration consistent with the extended conformation throughout the simulation. The unwinding process involves the breaking of hydrogen bonds at residues 74-81 [90]. These authors observe rigid motions of two domains around a single “hinge point” located here. Furthermore, pH titration experiments on CaM dimethylated with [¹³C] formaldehyde imply that the pK_a of Lys-75 is highly sensitive to the environmental changes such as peptide binding, indicating that the helical linker region unravels around this point [50]. Proteolysis of trypsin sensitive bonds lead to cleavage in Arg-74, Lys-75, and Lys-77 of the central helix which is not eliminated at high Ca²⁺ concentrations, while at intermediate concentrations there is an order of magnitude increase in the rate of proteolysis

indicating enhanced flexibility [127]. This behavior suggests that the linker may take on different roles depending on the solution conditions.

Perhaps equally important to simultaneously inducing bending and twisting motions by perturbing a single residue is the direction of the perturbation. All perturbations that give large overlaps with the targets fall along this line of perturbation within $\pm 10^\circ$, making use of the less crowded region between this and helix A (residues 5–19). Although the region has low solvent accessibility due to the presence of side chains, this direction is nevertheless a convenient pathway for proton uptake/release. Huang and Cheung have studied in detail the effect of H^+ and Ca^{2+} concentrations on activation of enzymes by CaM [128, 129]. Their findings suggest that the addition of Ca^{2+} exposes an amphipathic domain on CaM, whereas H^+ exposes a complimentary CaM binding domain on the target enzyme. The additional flux of H^+ might originate from CaM upon Ca^{2+} binding, or from transient cell conditions or both. Their findings also suggest that such changes might occur via subtle pH changes in the range of 6.9–7.5.

Ca^{2+} ion in EF-hand loop I in CaM is modulated by three Asp and one Glu residue (Figure 27.b). In the absence of a direct electrostatic interaction between E31 and Ca^{2+} ion, the side chain is expected to flip towards this gap, producing a local perturbation. A local scan of the possible isomeric states of the side chain of Glu31 yield only two possible conformations that will fit the gap; these alternative conformations are also shown on Figure 27 (b) as transparent traces.

Original backbone dynamics measurements made by ^{15}N NMR on Ca^{2+} -CaM indicated that the motion of the N- and C-terminal domains are independent [54]. Therein, a very high degree of mobility for the linker residues 78–81 is reported. These authors claim that their experiments support the idea the central linker acts merely as a flexible tether that keeps the two domains in close proximity. However, under different conditions, the two domains may well be communicating through the conformations assumed by the linker.

It is shown that helix A is stabilized upon calcium binding has been determined by frequency domain anisotropy measurements on unloaded and loaded CaM [130]. In the proposed model stemming from NMR analysis of a tetracysteine binding motif that has been engineered into helix A by site directed mutagenesis followed by fluorescent labeling, secondary structural changes in the linker orchestrate the release of helix A to allow for further Ca^{2+} binding upon activation of the C-terminal domain. Results show

that the large amplitude, nanosecond time scale motions occurring in this region are suppressed by Ca^{2+} loading to the N-terminal domain. These results are also corroborated by binding kinetics studies on fluorescently labeled samples with various degrees of Ca^{2+} loading [116]. Conversely, one may consider that helix A is destabilized once E31 side chain flips to release its grip on the Ca^{2+} ion. We identify pH changes as a possible source for such local conformational changes.

Our results reveal the signalling role of E31 in communicating information between the NTD and CTD of CaM. First, the mere protonation of this site (E31^+ run) is strong enough to reverse the effect of decreasing the IS, and localizing the conformations sampled to the region around the initial 3CLN structure (compare figures 9 and 16). Furthermore, by mutating it to a hydrophobic residue (E31A run), a new channel of conformational change whereby the linker assumes the frequently observed kinked conformation is made possible.



Figure 29 . Comparison of the structure attained in 3CLN* run 2 with the best-fitting NMR compatible one from the PDB file 2K0E. with 4.6 Å RMSD of the overall backbone, 2.5 Å in the N-lobe and 3.7 Å in the C-lobe.

We emphasize that although closed conformations of CaM are observed in both the 3CLN* and low IS 3CLN run, these have a straight linker conformation that is compatible with some of the NMR ensemble structures (Figure 29). The more frequently observed kinked conformation is only attainable after establishing the E31A mutation.

The bent conformation of the helix requires surmounting a high energy barrier under physiological conditions. In the NMR ensemble, 106 structures have bent

conformations ($l_{linker} < 30 \text{ \AA}$), 43 are straight ($l_{linker} > 32 \text{ \AA}$) and 11 are kinked but not bent $30 < l_{linker} < 32 \text{ \AA}$). We find that for bent linkers, the curve is always in the same direction, and occurs at residues 79-81. There is no general way of classifying the bent linker as stabilized by a salt bridge on either side of the kink (exceptions are rare; e.g. in one of the structures where the linker length is 22.9 \AA , there is a salt bridge between K75-E84; in another one of length 29.8 \AA , it occurs between K75-E82). Thus, the direct shifts between the bent-straight conformations must be prevented by repulsions between the NTD and CTD. We confirm that this is the case by the SMD simulations. By pulling in the correct direction, it is possible to achieve a kinked linker. However, since the relative positions of NTD and CTD are not optimal, the motion cannot be completed.

That the changes in E31 propagate to the CTD is corroborated by the relative helix orientations. In the wild type runs, the fluctuations in the relative orientation of the helices connecting the linker to the EF-hand IV domain (helix V-VII) is large (Table 6). On the other hand, in all the E31A runs, these fluctuations are greatly reduced to less than half and the region from the linker helix V all the way to the loop connecting EF-hand III to EF-hand IV is rigidified. MM results also corroborate the fact that structural manipulations between the open and closed conformations are best induced by displacements in two distinct regions: (i) Residues 17 and 18 in direct contact with E31 and at the controlling point of the angle helix I makes with helix II lead to large overlaps between the two conformations. It is with the disruption of this right angle positioning of the two helices that induces conformational changes in all the MD simulations. (ii) Manipulating residues 110, 115 and 117 which reside on the disordered loop that connects helix VI on EF-hand III to helix VII on EF-hand IV are also very effective. We note that, despite the disordered nature of this loop, the relative positioning of helices V-VI is fixed throughout all simulations, irrespective of the changes in condition.

By inspecting the MD trajectories, we find that the main difference between E31⁺ and E31A runs is that while the calcium binding motif is not disrupted in the former, E31 can no longer participate in the motif in the latter initiating the cascade of events that eventually trigger the large conformational change with the bent linker structure. The two communicating regions is shown in Figure 30.

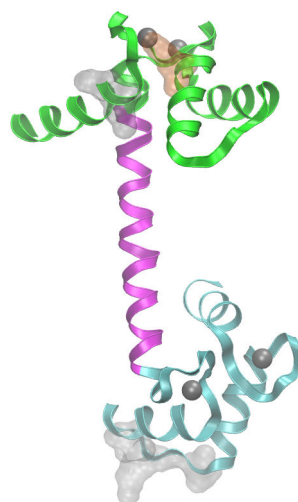


Figure 30. 3D representation of 3cln where the residues identified shown in surface representation for MM in silver and for PRS in orange on the protein.

4.3. Role of Electrostatics on the Conformational Change of CaM

The Ca^{2+} -CaM structure has to perform a simultaneous twisting and bending motion of the two lobes to overcome the energy barrier between the open and closed structures, exemplified by the 3cln and 1prw x-ray structures. However, the transition from the extended Ca^{2+} -CaM structure to the one with a bent linker is not likely to occur, unless there is a ligand clamped by the two lobes screening the interaction of the two domains (see, for example, CaM-myosin light chain kinase complex [29]). Also, it is known from Table 3 that many of the liganded structures of Ca^{2+} -CaM have low RMSD with 1prw; for example, 4.2, 3.6, 5.2, and 4.6 Å rmsd with 1lin, 1qiw, 2bbm and 1cdl, respectively. Thus, although 3cln dynamically has the propensity to complete the conformational change, the motion is prevented due to the large electrostatic repulsions between the two lobes which are predominantly negatively charged. However, in Ca^{2+} -CaM, as opposed to apo-CaM, there is also a partially positively charged region on the N-lobe [131]. Yet, the extended structure must perform the suitable motions to bring together this patch with the highly negatively charged C-lobe. That the interdomain PREs in Ca^{2+} -CaM depend on salt concentration lends support to the role of electrostatics on compaction [131].

Protonation of the ten residues in the 3CLN* system does not lower the barrier to the compact structure with bent linker, but creates an alternative path for the closing of Ca^{2+} -CaM. We calculate the interaction energy between the N- and the C-lobes using

the NAMD energy plugin in VMD (Figure 31). The electrostatic interactions are observed to make the major contribution to the nonbonded energy while the van der Waals part is observed to be an order of magnitude smaller. A dip in the energy occurs around 40 ns followed by two large dips at 67 and 120 ns. After 120 ns, the lobe-lobe energy enters an increasing trend for ca. 20 ns, and stabilizes thereafter. To associate the main events taking place during the conformational change with the interlobe energetics, we study the interactions between residues in the protein throughout the simulation. The first key interaction that leads to the permanent shrinkage of the distance between the two lobes is a salt bridge between the N-lobe and the linker (E7 – R74) formed permanently at 40 ns. We attribute the first significant reduction in the lobe-lobe interaction energy to the stabilizing effect of this salt bridge which also restricts the motions of the N-lobe.

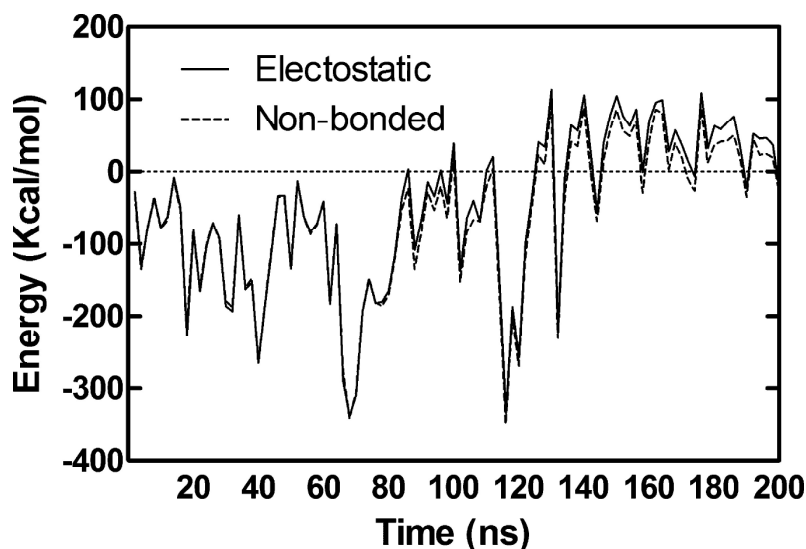


Figure 31. Nonbonded interaction energy between NTD and CTD in the 3CLN* MD simulation trajectory. Most contributions come from the electrostatic component (dashed line), displaying that the van der Waals term has negligible effect on the conformational change.

The transition state occurs at ca. 55 ns [18] with 20° swinging in the virtual torsional angle φ within 100 ps. This event changes the relative placement of the two lobes from that of opposite directions with respect to the linker to the same side. At 67 ns, the calcium ion in the third EF-hand motif partially dissociates. This region, located in the C-lobe, immediately following the linker, partially unfolds after this event and lets the C-lobe approach the N-lobe. The major cause of enhanced C-lobe dynamics is that the protonated E104, which is part of the EF-hand motif III, weakens the grip of the

motif on the Ca^{2+} . This is also the reason why the C-lobe attains a higher RMSD value than both the linker and the N-lobe from this point onward (recall Figure 12). The overall result of the loosened motif is the association of the C-lobe with the N-lobe which causes another dip in the energy. We note, however, that none of the Ca^{2+} completely dissociates from their EF-hand motifs in any portion of the simulation.

By 80 ns, the structure is further stabilized by the formation of a hydrophobic patch at the NTD–CTD interface. This event is accompanied by the raise in the interaction energy in the time range of 67–80 ns. Meanwhile, the CTD loosened at the EF-hand motif III performs a conformational search for a suitable region to latch on the rigidified NTD–linker. At 120 ns, the calcium ion of the EF-hand motif IV is shared between the two lobes, held in place by two negatively charged residues from the linker and the CTD, E84 and E139. However, this is a transient event, and in the last 40 ns, the structure attains its stabilized form.

We also monitor the repeatability of the structural change by an independent simulation for 100 ns and have observed a similar compaction of the overall structure (see Appendix in ref [18]). The conformational change in fact happens faster in that instance, and is again initiated by the formation of a salt bridge between the NTD and the linker. Furthermore, it appears that protonation of the ten residues, selected so as to mimic the most probable states of charged residues at pH 5, shifts the energy landscape such that the extended form is no longer preferred. A slightly more compact form is attained by first rigidifying the NTD–linker region by the formation of a salt bridge. While the linker keeps its helical structure, the EF-hand motif immediately following the linker operates as the conformationally flexible region searching for a suitable position to stabilize the structure by bringing together the partially positively charged region on the NTD with the highly negatively charged region on the CTD. However, it should be emphasized that these observations do not preclude other possible compaction events that may utilize the bending of the linker under different environmental conditions such as that observed in the E31A_R1.

While mimicing the lower pH conditions introduces a fast compaction event as described in detail above, the same sequence of events is also seen at IS low compared to physiological conditions (compare Figures 9 and 13). In fact, the three regions visited during the 3CLN low IS run and those in 3CLN* runs are very similar. The time scale of the events is relatively faster in the former compared to the latter (~80 ns vs. 200 ns), however the latter has a much more negatively charged surface. Thus, in both of these

runs, the compaction is achieved by the rotation of the two lobes in the torsional angle space followed by engaging the NTD and CTD without binding the linker at their appropriate positioning. These are plausible conformations, as confirmed by the presence of nearby structures in the NMR ensemble of structures.

Under physiological IS and pH (150 mM and 7.4, respectively), the time scale of jumps between open and compact conformers is on the order of 100 microseconds [34]. Thus, both lowering pH and/or IS accelerate the time scale of the events. In fact, we find that the conformation is anchored to the initial state in the 150 mM run (Figure 24). The torsional degrees of freedom that are active in the above situations are frozen in this case. This observation is again found its origin in electrostatics.

To quantify the distribution of charges around the protein, we have calculated the radial distribution function (Figure 32) between the heavy atoms of the protein and the Na^+ ions (as well as Cl^- ions for the 150 mM IS case). We find that the Na^+ ions approach the protein surface much less at low IS compared to high IS. The 3CLN* run (not shown) is intermediate between the two. The first peak belongs solely to the contact of Na^+ ions with O^- atoms of the negatively charged residues. The second peak is for the approach of the ions to all other heavy atoms. Interestingly, although there are plenty of positively charged residues on the surface of the protein, Cl^- ions (which only exist in high IS runs) rarely interact with them. When decomposed into the different regions on the protein (inset) we find that the most significant interaction is with the linker residues, followed by those of the C-lobe and even less so with the N-lobe at low IS. The trend is different at high IS, where the largest interactions with the Na^+ ions occur at the C-lobe.

As per Israelachvili [132], the interactions of solutes in water have an oscillatory hydration force component. At distances shorter than 3-4 nm (as is the case in this study) DLVO forces are no longer dominant. For smooth surfaces, the fluctuations are more pronounced while for rough surfaces the fluctuations are smeared out. When the two domains are relatively far – as in 3cln – it may be assumed that roughness is perceived to be relatively low by the two domains. Significant amount of anions and cations are introduced to the region permeating the domains when IS is increased. In this case, the fluctuation forces are expected to diminish, since there will be a decrease in solvation forces that occur due to ordering of the water molecules in between two surfaces.

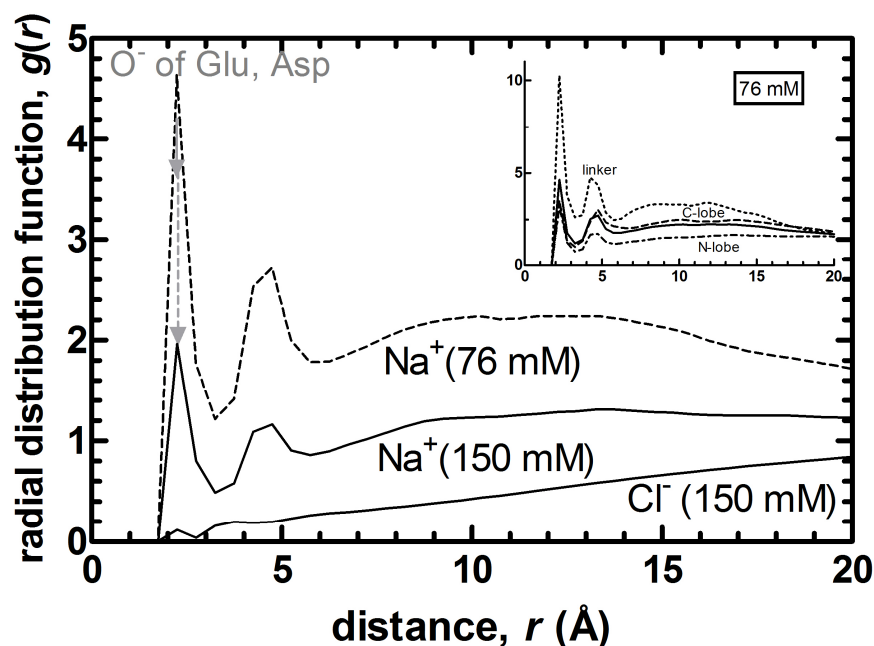


Figure 32. The radial distribution between the heavy atoms on 3cln and the ions Na^+ ions (as well as Cl^- ions for the 150 mM IS case) in the solvent for 76 mM IS and 150 mM IS conditions.

Increasing salt concentration alters the electrostatic interactions between CaM and target peptides. Then, favorable side chain interactions with target peptides become more probable, thus making it possible to bind to a target peptide and change conformation. Target molecule also provides an additional effect to the screening between the two strongly repulsive domains. In the absence of the target molecule, salt screening by itself is not sufficient to alter the energy landscape towards the compact structure. It does however, pave the way for the Ca^{2+} -CaM structure to enter the pathway leading to compact structure.

4.4. Comparison of the Local Perturbations Introduced by PRS and MM

In both the MM and PRS methodologies further developed in this work, the system responds in the linear response regime. The latter has its theoretical basis in the linear response theory [22]. For the former, we ensure that the perturbations introduced are in the size range giving linear response [35]. However, the two methods have fundamental differences that would be reflected in the thermodynamical functions that are optimized under the inserted perturbation.

In MM, the perturbation is in the form of a fixed displacement in a random direction, with the rest of the system (including water) being minimized and allowed to

adjust according to this local change. The displacement in the MM protocol generates a virtual force distributed over all residues; this force distribution in turn leads to the displacements recorded in the whole system. In PRS, the perturbation is in the form of a force introduced on one residue in a given direction, while the whole system is displaced in response to this single residue perturbation.

In PRS, the response of the system is strongly dependent on the covariance matrix that carries the information of the cross correlations between the residues, obtained from an MD trajectory. Furthermore, the covariance matrix is constructed within a selected time window [106]. Inasmuch as one selects the window size that samples a single conformational substate, the results will reflect the effective interactions in the system. In the highly charged protein system that we have in CaM, these are expected to be mostly of electrostatic origin. Nevertheless, since all interactions, including solvation, are taken into account within a given well-sampled substate, the recorded response reflects free energetic differences. We note that once the cross correlations are obtained in the full atomic description, the response is calculated at a coarse grained level.

In contrast, MM is an all atom methodology taking advantage of the force field operative at the atomic scale. However, the perturbations are local and therefore the harder degrees of freedom such as distorted bond lengths and angles, as well as van der Waals forces between overlapping atoms will take precedence in the response of the system. The solvation contribution is minimal with only small variations in a fixed water arrangement in the vicinal layer. Thus, the recorded response is mainly energetic.

Yet, we find that both approaches designate the same two regions on the protein structure important for inserting perturbations to get the expected conformational change. These are (i) the region around the Ca^{2+} atom in EF-hand loop I in the NTD and (ii) the loop joining EF-hand loops III and IV in the CTD. The actual identity of the residues obtained within these regions are different however. PRS implicates charged residues (E31, D122) while MM finds polar and hydrophobic residues (L18, F19, T110, K115, T117). These results reflect the differences inherent in the thermodynamic functions optimized by the two approaches.

CHAPTER 5

Conclusion and Future Work

5.1. Major Findings and Contribution

Conformational multiplicity of CaM has been the subject of numerous studies (see, for example, refs 47 and 48 and references cited therein). Those states populated by CaM under different environmental conditions enhance our understanding of the role of conformational selection and induced fit models in its activation. We have studied the manipulation of the extended structure of Ca^{2+} loaded CaM to seven different structures reported in literature by using simple perturbation-response approaches. The methods used in this work provide strong tools for following residue interactions in a systematic and well-defined manner. Residue fluctuations and displacement vector directions reveal many interesting features of phenomena underlying interactions in proteins.

Our main findings from the local perturbations introduced by PRS and MM indicate the following: (i) Reduction of the CS to a few degrees of freedom conveniently describe the main features of the conformational changes of Ca^{2+} -CaM. These are represented via a simultaneous twist and bend motion of the two lobes with respect to each other. (ii) For five of the seven structures, the conformational change occurs as independent of ligand size, although the RMSD values may be large. (iii) The vector set describing the conformational change, however, is not simply described by a single underlying collective mode, but corresponds to some motion that seems to be stimulated by perturbing a particular residue (E31) in a particular direction. Thus, the CaM structure is optimized for proper signals that influence E31, and not for responding to random stimuli. (iv) The perturbations of E31 induces coupled counter twisting and bending motions in the linker, and the bend is induced around residues found susceptible to dynamical changes via NMR and proteolysis experiments. (v) Independently, we find E31 to also be a unique residue (out of a total of 52 charged ones) whose ionization state is sensitive to subtle pH variations in the physiological

range as corroborated with four different pK_a calculation approaches. The transition between charged/uncharged states in E31 occurs in a narrow pH window of ~ 6.5 – 7.5 . Combined with item (iv) above, E31 is thus implicated to be a center for conformation control via differential pH gradients. (vi) MM reveals residues 17 and 117 as the conformational switch initiating residues. Our overall analysis implicates a communication pathway between these regions on the two lobes, mediated by E31.

We also studied the role of electrostatics in altering the conformation distributions as well as the dynamics of Ca^{2+} –CaM using extensive MD simulations starting from its extended and compact forms under standard protonation states (3CLN and 1PRW, respectively) as well as by protonating ten residues with relatively high pK_a values (3CLN*) and by protonating only residue E31⁺. In 1PRW run, the N-lobe, C-lobe and the linker maintain their local stability. For 3CLN, the linker stays in the extended conformation where as the NTD and CTD relaxes to a conformation that is also found in the 3CLN* run. As confirmed by RMSD profiles, and the relative orientations of the lobes quantified by $(l_{\text{linker}}, \phi)$ maps, these structures are energetically stable and sample conformations in a local energy minimum. Altering the protonation state of residues which are highly sensitive to pH (3CLN*) or mutating only E31 to Alanine (E31A) readily lead to conformational switching to more compact forms on sub-100 ns time scales. The structures are compatible with the FRET results both in terms of the interlobe distances and the fact that lowering the pH to 5.0 leads to a single stable state as opposed to the presence of at least two distinct forms at 7.4 pH [48]. The final, equilibrated structure is also similar to several ones from the ensemble of conformations representing the structure and dynamics of Ca^{2+} –CaM as related to NMR derived parameters.

Our current view suggests that at high pH, barrier crossing to the compact form is prevented by the repulsive electrostatic interactions between the two lobes at physiological pH. Because of the remote location of the oppositely charged patches on the two domains, the conformational change only occurs when a simultaneous torsion and bending of the two lobes with respect to each other takes place, a rare event occurring on 100 μs time scales at physiological conditions. Lowering the pH significantly alters the charge distribution on the protein surface, leading to a change in the overall conformational landscape. Thus, even at elevated pH, the protein might ease the passage between the different conformers by acquiring protons from the solvent environment. It is interesting to note that these residues with elevated pK_a values are

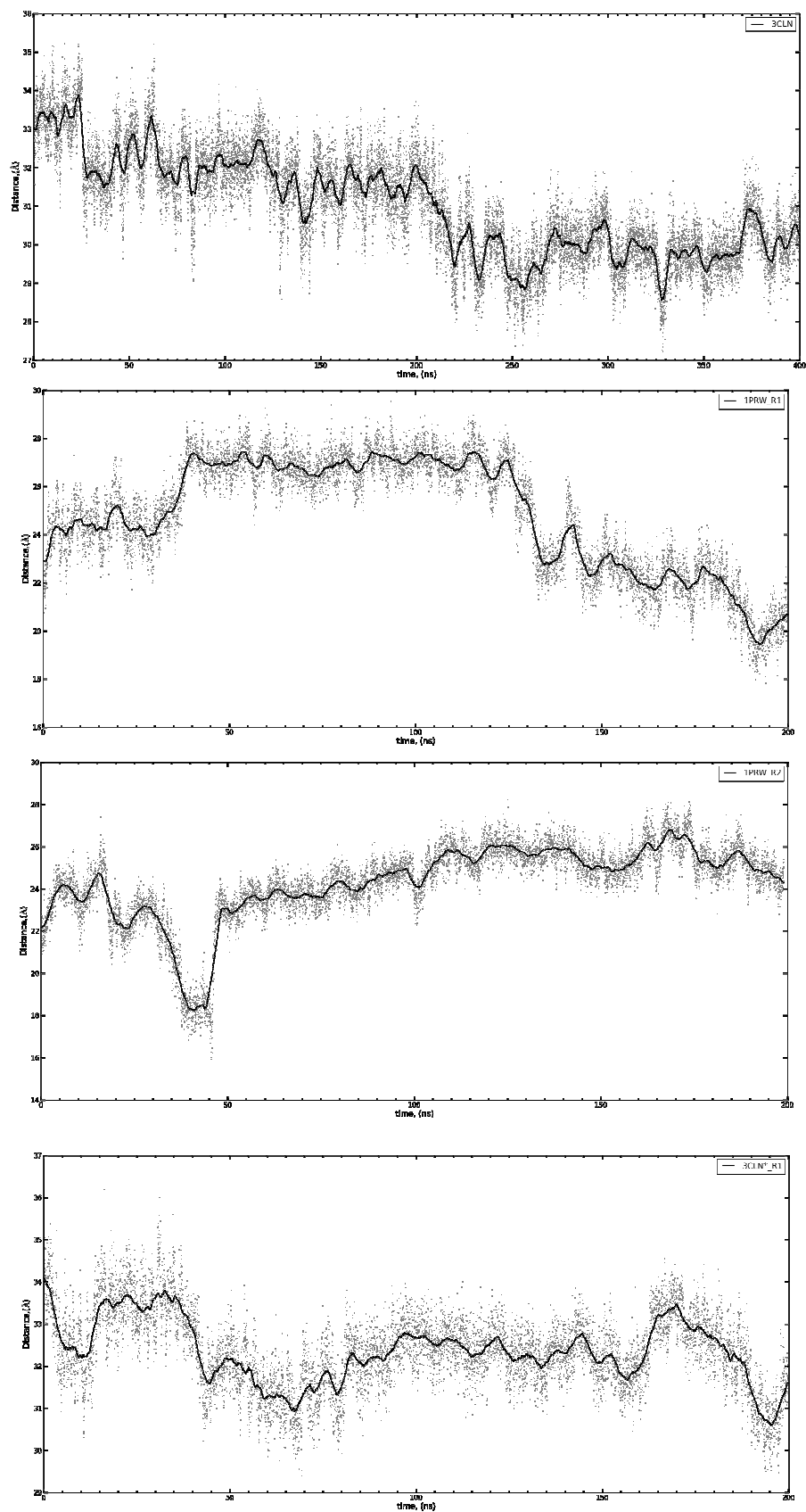
also strategically located at regions of the protein which may mechanically induce allosteric changes (E31 and D122 are in close contact with residues 17 and 117 found in the MM analysis. They are also directly implicated in PRS analysis). Adaptation to different pH values in various subcellular compartments is thought to be directly related to protein stability and pH at optimal binding affinity for interacting proteins.

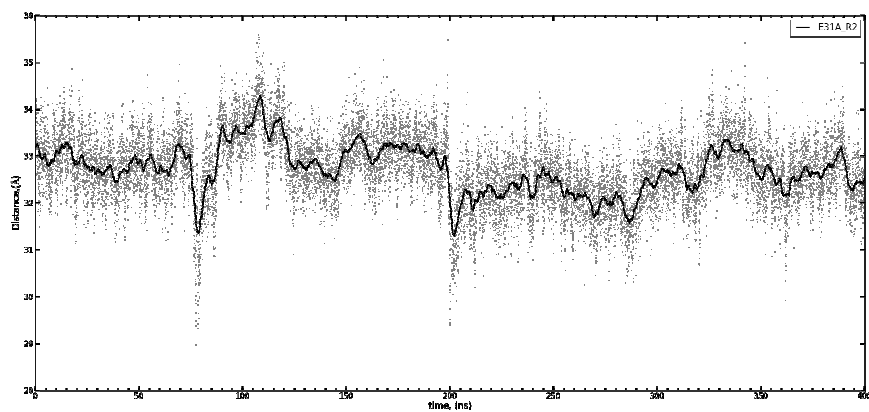
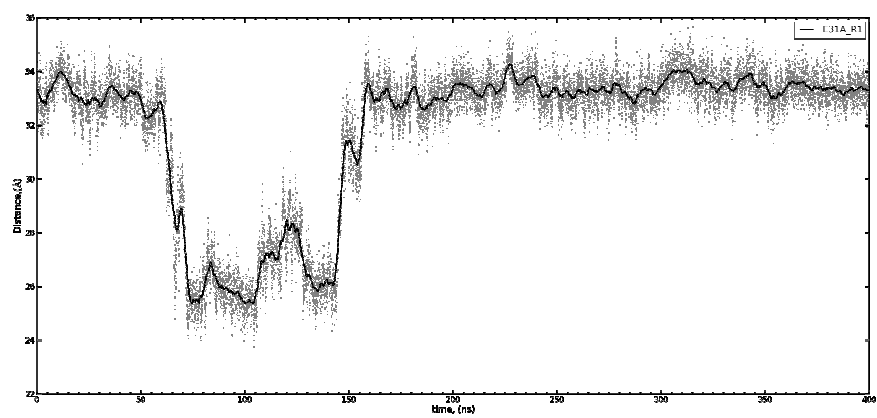
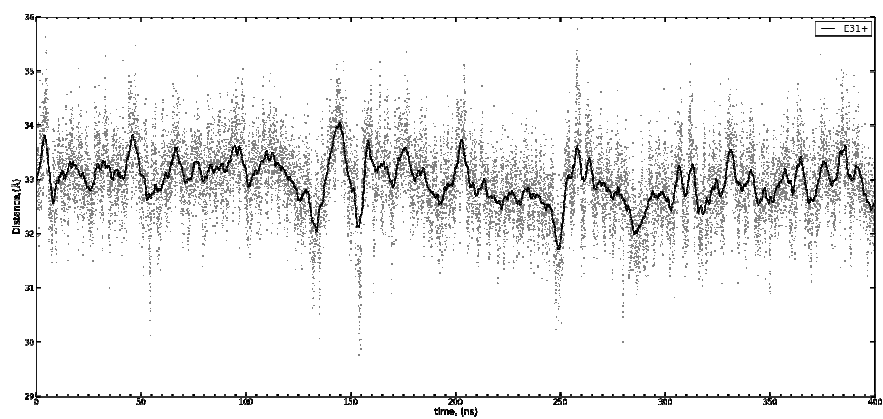
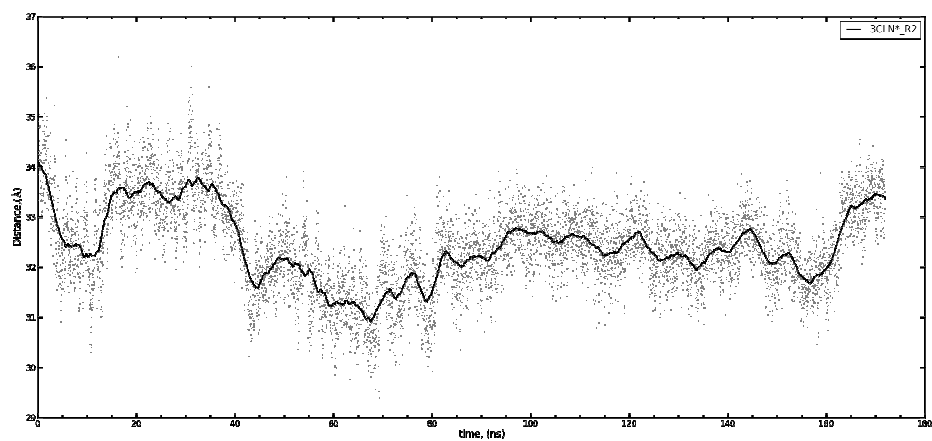
Also, our previous study of FBP via PRS also indicated charged allosteric residues to coordinate ion release, again implicating a coupled electrostatic–mechanical effect. In FBP, the remarkably high association constants on the order of 10^{17} – 10^{22} M^{-1} not only suggest that it is fairly easy for FBP to capture the ion, but also poses the question of how it is released once transported across the periplasm. Allosteric control using the different electrostatic environments in physiological conditions was put forth as a possible operating mechanism. It is remarkable that residues predicted by PRS that reproduce the targeted conformational change are predominantly charged (E6, K30, E31), although the electrostatics is not directly implemented in the PRS technique whereas the ones that are predicted by MM are uncharged residues (S17, T117).

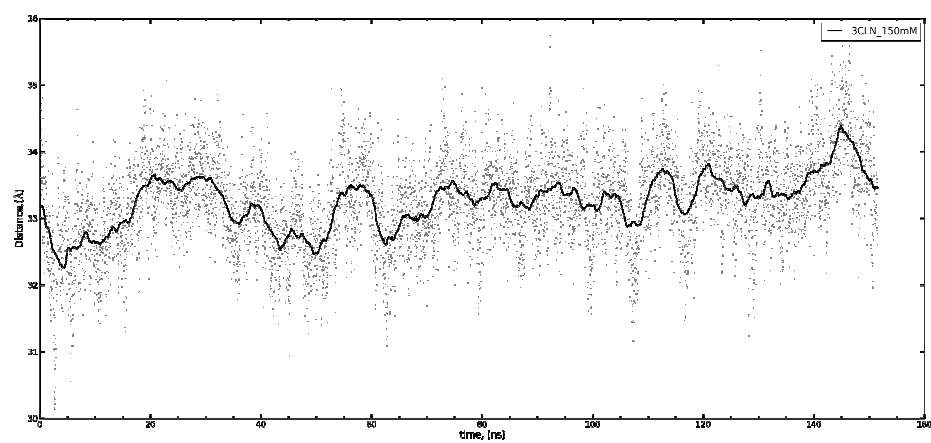
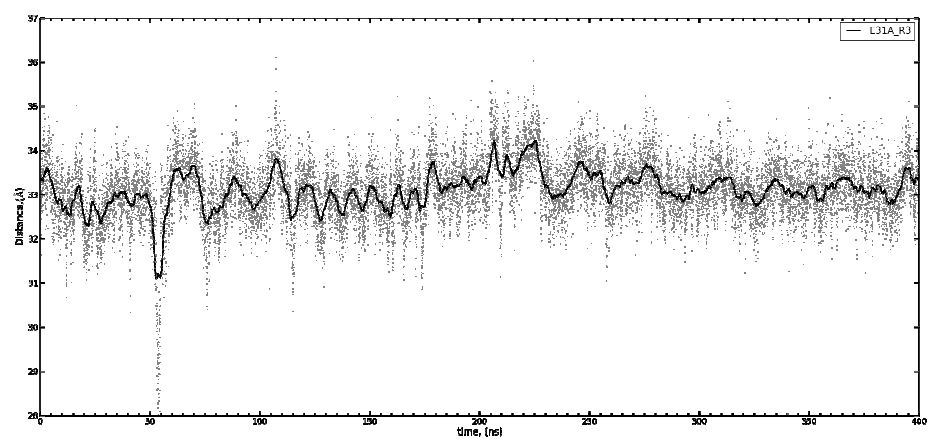
5.2. Future Work

In addition to NMR, x-ray, FRET, and single molecule force spectroscopy methods, combined ion mobility/mass spectroscopy methods are also becoming attractive for investigating the effect of different environments on conformation distributions of proteins. A recent example is electro-spraying experiments on various CaM structures which indicate that the extended conformation is abundant at higher charge states of the protein. Developing simple and efficient methods such as PRS and MM for investigating the relationship between modulated electrostatic environment of the protein and its mechanical response will therefore continue to be an important area of research, particularly as a unique approach to attack the problem of identifying adaptations of proteins to subcellular pH.

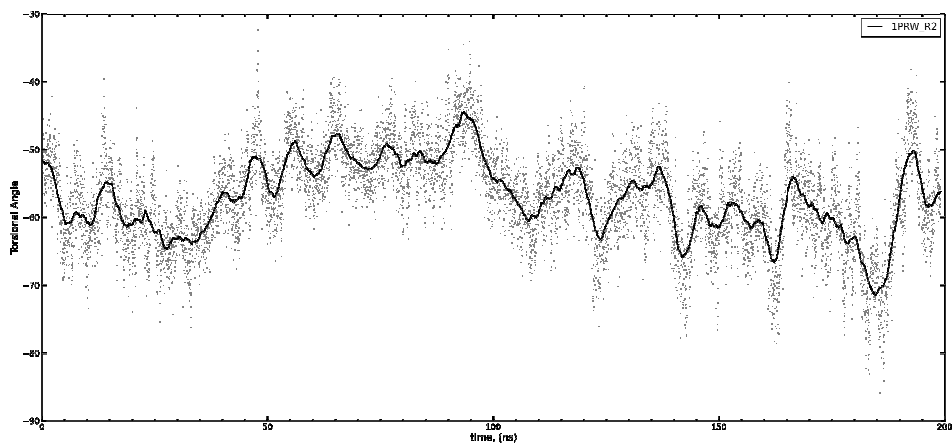
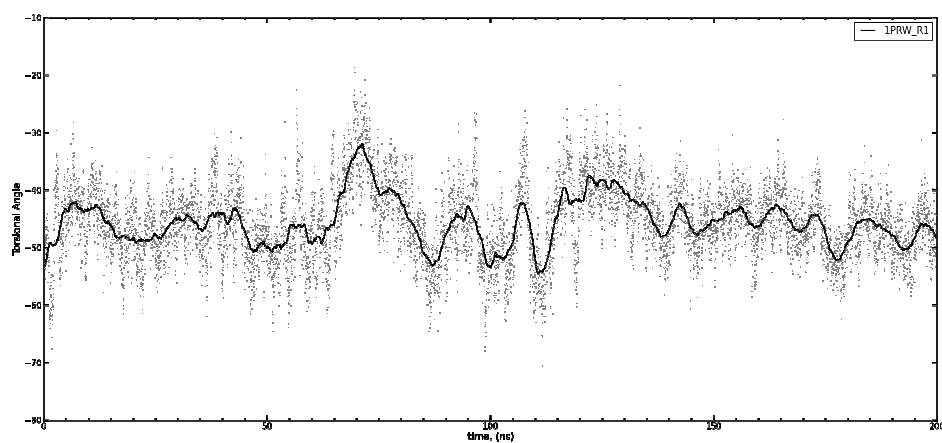
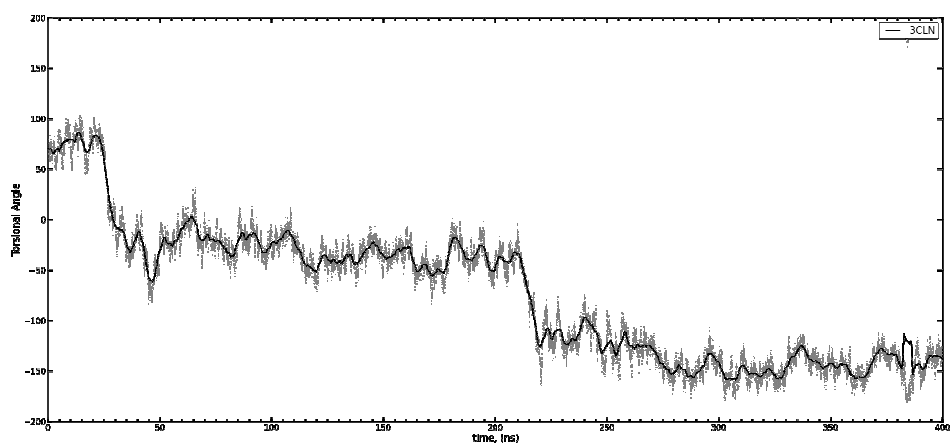
Appendix A. Time Evolution of End-to-End Distance in MD trajectories

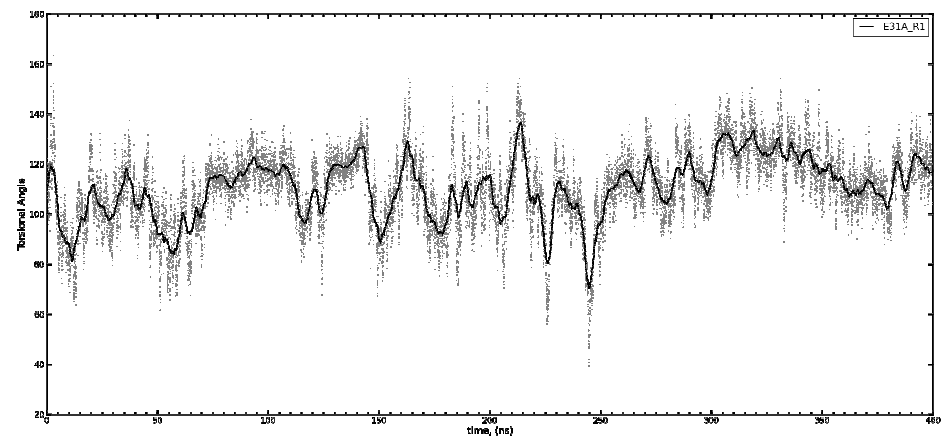
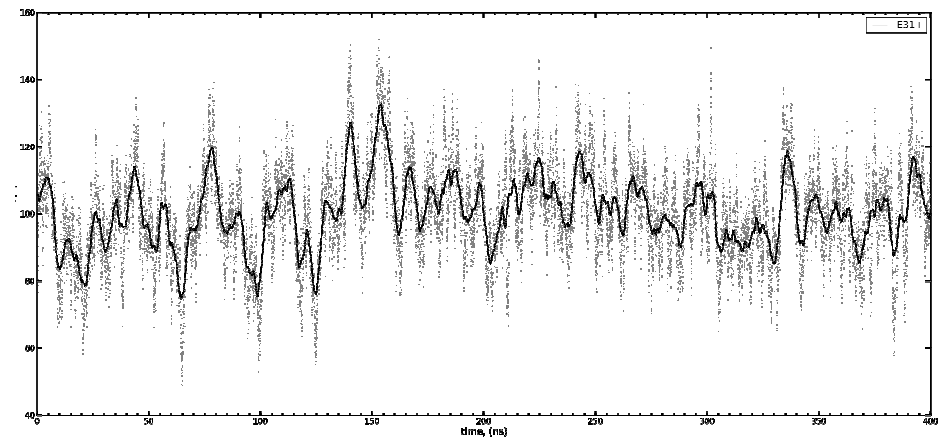
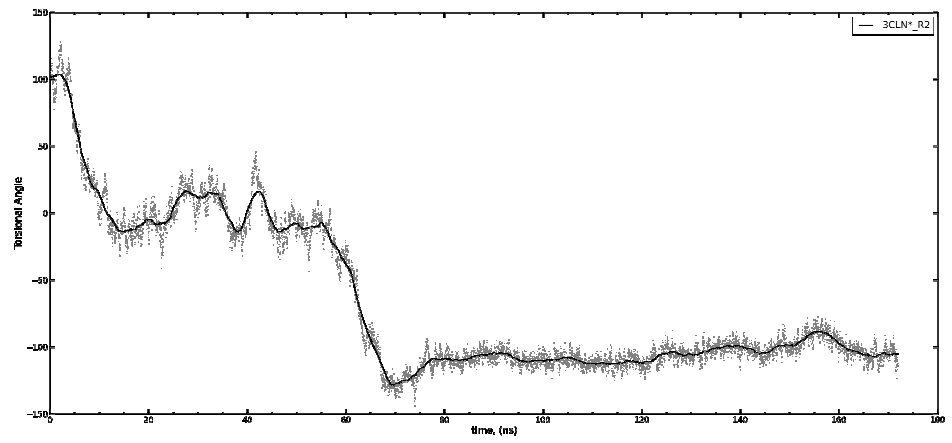
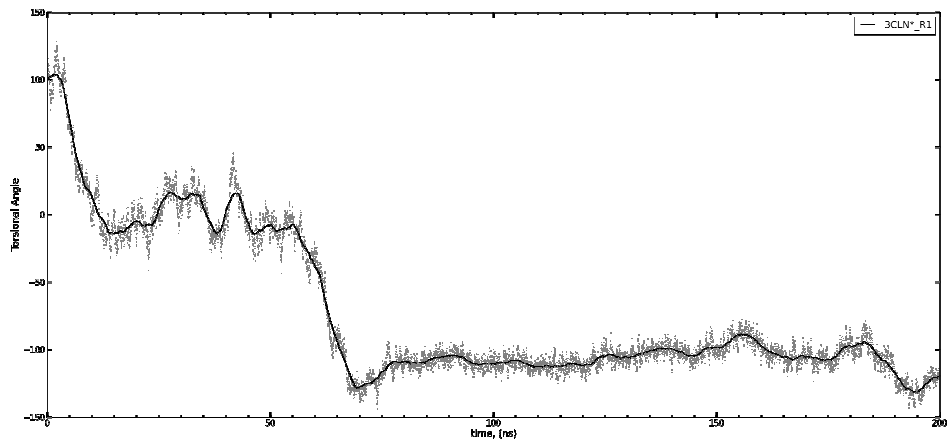


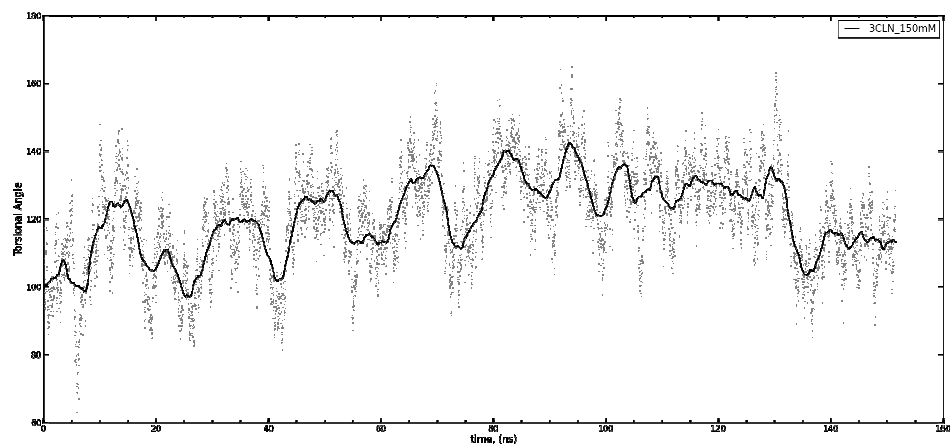
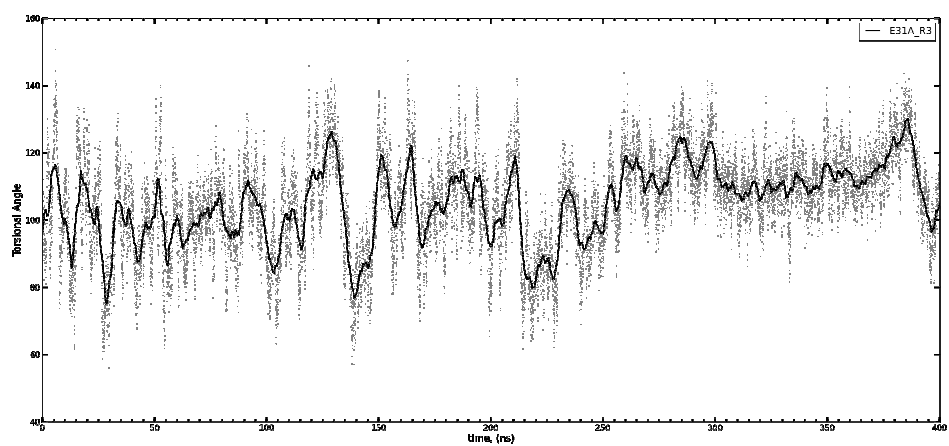
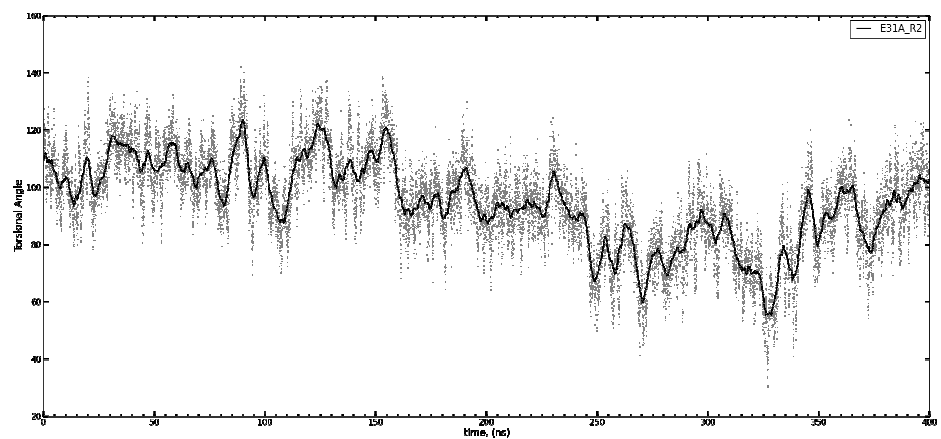




Appendix B. Time Evolution of Torsional Angle in the MD trajectories







Appendix C. Molecular Mechanics Script

This script is written in tcl programming language and operated in VMD tcl console using the command below:

```
vmd -dispdev text -eofexit <name.tcl> log.txt &
```

This command operates vmd in text mode and runs the tcl script without opening VMD with display and log.txt file contains the information that will be written on the display.

The following lines belong to the name.tcl script.

```
#load the input structure and input minimized coordinate files.
mol load psf inputstructure.psf pdb inputstructure_minimized.pdb
#specify the radius of the sphere.
set radius 0.5
#specify the number of loops with variable i starting from 1 to selected loop number.
for {set i 1} {$i<I} {incr i} {
#loop over the residue number of the protein with variable r ; starting number can change
according to the starting residue number in the pdb file. Each time perturbation will be given to
the selected residue Cα
for {set r 1} {$r<R} {incr r} {
set a [atomselect top "resid $r and name CA"]
set b [atomselect top all]
#calculate a random displacement vector in x,y,z directions.
set c [expr (2*rand()-1)]
set d [expr (2*rand()-1)]
set e [expr (2*rand()-1)]
#calculate the magnitude of the displacement vector.
set f [expr $c**2+$d**2+$e**2]
set mag [expr {sqrt($f)}]
#divide the radius of the sphere by the magnitude of the displacement vector.
set h [expr ($radius/$mag)]
#multiply each direction with this constant.
set cc [expr $h*$c]
set dd [expr $h*$d]
set ee [expr $h*$e]
#get the updated displacement vector x,y,z directions.
set x [list $cc]
set y [list $dd]
set z [list $ee]
#perturb the selected residue with the given direction.
$a moveby "$x $y $z"
$a get {x y z}
#set occupancy of the perturbed residue Cα to 1 in order to fix this atom in space.
$b set {occupancy} {0}
$a set {occupancy} {1} 1
#write the coordinates of all the system to the specified pdb file
$b writepdb $i.res.$r.pdb
#move the coordinates back to their initial values
set ccc [expr -1 * $cc]
set ddd [expr -1 * $dd]
set eee [expr -1 * $ee]
```

```
set cccc [list $ccc]
set dddd [list $ddd]
set eeee [list $eee]
$a moveby "$ccc $ddd $eee"
```

```
#perform minimization by using the perturbed coordinates.
exec env NAMD_psf=3CLN_20ns1stfrm_10Awb.psf\
NAMD_pdb=$i.res.$r.pdb\
output=$i.res.$r.\
#this is a standard NAMD running command given for a cluster with 48 cpus with charmrun
/home_palamut2/catilgan/programs/NAMD_2.8_Linux-x86_64/charmrun +p48
/home_palamut2/catilgan/programs/NAMD_2.8_Linux-x86_64/namd2 mm.conf > $i.res.$r.txt
}
}
```

References

1. Garcia-Moreno, B., *Adaptations of proteins to cellular and subcellular pH*. Journal of Biology (London), 2009. **8**: p. 98.
2. Berger-Wolf, T.Y., T.M. Przytycka, and M. Singh, *Dynamics of Biological Networks*. Pacific Symposium on Biocomputing 2009, ed. R.B. Altman, et al. 2009. 174-177.
3. Przytycka, T.M., M. Singh, and D.K. Slonim, *Toward the dynamic interactome: it's about time*. Briefings in Bioinformatics, 2010. **11**(1): p. 15-29.
4. Ming, D.M. and M.E. Wall, *Interactions in native binding sites cause a large change in protein dynamics*. Journal of Molecular Biology, 2006. **358**(1): p. 213-223.
5. Boehr, D.D., et al., *The dynamic energy landscape of dihydrofolate reductase catalysis*. Science, 2006. **313**(5793): p. 1638-1642.
6. Bui, J.M. and J.A. McCammon, *Protein complex formation by acetylcholinesterase and the neurotoxin fasciculin-2 appears to involve an induced-fit mechanism*. Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(42): p. 15451-15456.
7. Grant, B.J., A.A. Gorfe, and J.A. McCammon, *Large conformational changes in proteins: signaling and other functions*. Current Opinion in Structural Biology, 2010. **20**(2): p. 142-147.
8. Marsh, J.A., S.A. Teichmann, and J.D. Forman-Kay, *Probing the diverse landscape of protein flexibility and binding*. Current Opinion in Structural Biology, 2012. **22**(5): p. 643-650.
9. Mueller, T.D. and J. Nickel, *Promiscuity and specificity in BMP receptor activation*. Febs Letters, 2012. **586**(14): p. 1846-1859.
10. Tzeng, S.-R. and C.G. Kalodimos, *Protein dynamics and allostery: an NMR view*. Current Opinion in Structural Biology, 2011. **21**(1): p. 62-67.
11. Gerstein, M., A.M. Lesk, and C. Chothia, *Structural Mechanisms for Domain Movements in Proteins*. Biochemistry, 1994. **33**(22): p. 6739-6749.
12. Koshland, D.E.J., Nemethy, G., and Filmer, G., *Comparison of experimental binding data and theoretical models in proteins containing subunits*. Biochemistry, 1966. **5**: p. 365-385.
13. McCammon, J.A., Gelin, B.R., Karplus, M., and Wolynes, P.G., *The hinge-bending motion in lysozyme*. Nature, 1976. **262**: p. 325-326.
14. Monod, J., Wyman, J., and Changeux, J.P. , *On the nature of allosteric transitions: a plausible model* Journal of Molecular Biology 1965. **12**: p. 88-118.
15. Perutz, M.F., et al., *The stereochemical mechanism of the cooperative effects in hemoglobin revisited*. Annual Review of Biophysics and Biomolecular Structure, 1998. **27**: p. 1-34.
16. Bertrand, G.-M., *Adaptations of proteins to cellular and subcellular pH*. Journal of Biology, 2009. **8**(11): p. 98.
17. Atilgan, A.R., A.O. Aykut, and C. Atilgan, *Subtle pH differences trigger single residue motions for moderating conformations of calmodulin*. Journal of Chemical Physics, 2011. **135**(15).
18. Negi, S., et al., *Calmodulin Readily Switches Conformation upon Protonating High pK(a) Acidic Residues*. Journal of Physical Chemistry B, 2012. **116**(24): p. 7145-7153.

19. Chin, D. and A.R. Means, *Calmodulin: a prototypical calcium sensor*. Trends in Cell Biology, 2000. **10**(8): p. 322-328.
20. Ikura, M. and J.B. Ames, *Genetic polymorphism and protein conformational plasticity in the calmodulin superfamily: Two ways to promote multifunctionality*. Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(5): p. 1159-1164.
21. Junker, J.P. and M. Rief, *Single-molecule force spectroscopy distinguishes target binding modes of calmodulin*. Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(34): p. 14361-14366.
22. Atilgan, C. and A.R. Atilgan, *Perturbation-Response Scanning Reveals Ligand Entry-Exit Mechanisms of Ferric Binding Protein*. Plos Computational Biology, 2009. **5**(10).
23. Atilgan, C., et al., *Manipulation of Conformational Change in Proteins by Single-Residue Perturbations*. Biophysical Journal, 2010. **99**(3): p. 933-943.
24. Flores, S., et al., *The database of macromolecular motions: new features added at the decade mark*. Nucleic Acids Research, 2006. **34**: p. D296-D301.
25. Keskin, O., *Binding induced conformational changes of proteins correlate with their intrinsic fluctuations: a case study of antibodies*. BMC Structural Biology, 2007. **7**.
26. Babu, Y.S., C.E. Bugg, and W.J. Cook, *Structure of calmodulin refined at 2.2 Å resolution*. Journal of Molecular Biology, 1988. **204**(1): p. 191-204.
27. Fallon, J.L. and F.A. Quijcho, *A closed compact structure of native Ca²⁺-calmodulin*. Structure, 2003. **11**(10): p. 1303-U7.
28. Harmat, V., et al., *A new potent calmodulin antagonist with arylalkylamine structure: Crystallographic, spectroscopic and functional studies*. Journal of Molecular Biology, 2000. **297**(3): p. 747-755.
29. Ikura, M., et al., *Solution structure of a calmodulin-target peptide complex by multidimensional NMR*. Science, 1992. **256**(5057): p. 632-638.
30. Meador, W.E., A.R. Means, and F.A. Quijcho, *Target enzyme recognition by calmodulin- 2.4-angstrom structure of a calmodulin-peptide complex*. Science, 1992. **257**(5074): p. 1251-1255.
31. Osawa, M., et al., *Solution structure of calmodulin-W-7 complex: The basis of diversity in molecular recognition*. Journal of Molecular Biology, 1998. **276**(1): p. 165-176.
32. Vandonselaar, M., et al., *Trifluoperazine-induced conformational change in Ca²⁺-calmodulin*. Nature Structural Biology, 1994. **1**(11): p. 795-801.
33. Yun, C.H., et al., *Structure of potato calmodulin PCM6: the first report of the three-dimensional structure of a plant calmodulin*. Acta Crystallographica Section D-Biological Crystallography, 2004. **60**: p. 1214-1219.
34. Price, E.S., M.S. DeVore, and C.K. Johnson, *Detecting Intramolecular Dynamics and Multiple Forster Resonance Energy Transfer States by Fluorescence Correlation Spectroscopy*. Journal of Physical Chemistry B, 2010. **114**(17): p. 5895-5902.
35. Baysal, C. and A.R. Atilgan, *Coordination topology and stability for the native and binding conformers of chymotrypsin inhibitor 2*. Proteins-Structure Function and Genetics, 2001. **45**(1): p. 62-70.
36. Ikeguchi, M., et al., *Protein structural change upon ligand binding: Linear response theory*. Physical Review Letters, 2005. **94**(7).
37. Gerstein, M. and W. Krebs, *A database of macromolecular motions*. Nucleic Acids Research, 1998. **26**(18): p. 4280-4290.

38. Di Russo, N.V., et al., *pH-Dependent Conformational Changes in Proteins and Their Effect on Experimental pK(a)s: The Case of Nitrophorin 4*. Plos Computational Biology, 2012. **8**(11).
39. Gagne, S.M., M.X. Li, and B.D. Sykes, *Mechanism of direct coupling between binding and induced structural change in regulatory calcium binding proteins*. Biochemistry, 1997. **36**(15): p. 4386-4392.
40. Gao, Z.H., et al., *Activation of 4 enzymes by 2 series of calmodulin mutants with point mutations in individual CA(2+) binding-sites*. Journal of Biological Chemistry, 1993. **268**(27): p. 20096-20104.
41. Maune, J.F., C.B. Klee, and K. Beckingham, *CA2+ binding and conformational change in 2 series of point mutations to the individual CA2+-binding sites of calmodulin*. Journal of Biological Chemistry, 1992. **267**(8): p. 5286-5295.
42. Pedigo, S. and M.A. Shea, *Quantitative endoproteinase gluc footprinting of cooperative CA2+ binding to calmodulin-proteolytic susceptibility of E31 and E87 indicates interdomain interactions*. Biochemistry, 1995. **34**(4): p. 1179-1196.
43. Shi, C., J.A. Wallace, and J.K. Shen, *Thermodynamic Coupling of Protonation and Conformational Equilibria in Proteins: Theory and Simulation*. Biophysical Journal, 2012. **102**(7): p. 1590-1597.
44. Carafoli, E., *Calcium signaling: A tale for all seasons*. Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(3): p. 1115-1122.
45. Wei, J.W., H.P. Morris, and R.A. Hickie, *Positive correlation between calmodulin content and hepatoma growth-rates*. Cancer Research, 1982. **42**(7): p. 2571-2574.
46. Johnson, C.K., *Calmodulin, conformational states, and calcium signaling. A single-molecule perspective*. Biochemistry, 2006. **45**(48): p. 14233-14246.
47. Slaughter, B.D., et al., *Single-molecule resonance energy transfer and fluorescence correlation spectroscopy of calmodulin in solution*. Journal of Physical Chemistry B, 2004. **108**(29): p. 10388-10397.
48. Slaughter, B.D., et al., *Conformational substates of calmodulin revealed by single-pair fluorescence resonance energy transfer: Influence of solution conditions and oxidative modification*. Biochemistry, 2005. **44**(10): p. 3694-3707.
49. Ikura, M., *Calcium binding and conformational response in EF-hand proteins*. Trends in Biochemical Sciences, 1996. **21**(1): p. 14-17.
50. Zhang, M., T. Tanaka, and M. Ikura, *Calcium-induced conformational transition revealed by the solution structure of apo calmodulin*. Nature Structural Biology, 1995. **2**(9): p. 758-767.
51. Bayley, P.M., W.A. Findlay, and S.R. Martin, *Target recognition by calmodulin: Dissecting the kinetics and affinity of interaction using short peptide sequences*. Protein Science, 1996. **5**(7): p. 1215-1228.
52. Crivici, A. and M. Ikura, *Molecular and structural basis of target recognition by calmodulin*. Annual Review of Biophysics and Biomolecular Structure, 1995. **24**: p. 85-116.
53. Chen, B.W., et al., *Dynamic motion of helix A in the amino-terminal domain of calmodulin is stabilized upon calcium activation*. Biochemistry, 2005. **44**(3): p. 905-914.

54. Barbato, G., et al., *Backbone dynamics of calmodulin studied by N-15 relaxation using inverse detected 2-dimensional NMR-spectroscopy - the central helix is flexible*. Biochemistry, 1992. **31**(23): p. 5269-5278.
55. Yamniuk, A.P. and H.J. Vogel, *Calmodulin's flexibility allows for promiscuity in its interactions with target proteins and peptides*. Molecular Biotechnology, 2004. **27**(1): p. 33-57.
56. Wilson, M.A. and A.T. Brunger, *The 1.0 angstrom crystal structure of Ca²⁺-bound calmodulin: an analysis of disorder and implications for functionally relevant plasticity*. Journal of Molecular Biology, 2000. **301**(5): p. 1237-1256.
57. Bertini, I., et al., *Experimentally exploring the conformational space sampled by domain reorientation in calmodulin*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(18): p. 6841-6846.
58. Hoeflich, K.P. and M. Ikura, *Calmodulin in action: Diversity in target recognition and activation mechanisms*. Cell, 2002. **108**(6): p. 739-742.
59. Vetter, S.W. and E. Leclerc, *Novel aspects of calmodulin target recognition and activation*. European Journal of Biochemistry, 2003. **270**(3): p. 404-414.
60. Schumacher, M.A., et al., *Structure of the gating domain of a Ca²⁺-activated K⁺ channel complexed with Ca²⁺/calmodulin*. Nature, 2001. **410**(6832): p. 1120-1124.
61. Drum, C.L., et al., *Structural basis for the activation of anthrax adenyl cyclase exotoxin by calmodulin*. Nature, 2002. **415**(6870): p. 396-402.
62. Bertini, I., et al., *Conformational Space of Flexible Biological Macromolecules from Average Data*. Journal of the American Chemical Society, 2010. **132**(38): p. 13553-13558.
63. Vendruscolo, M. and C.M. Dobson, *Dynamic visions of enzymatic reactions*. Science, 2006. **313**(5793): p. 1586-1587.
64. Kern, D. and E.R.P. Zuiderweg, *The role of dynamics in allosteric regulation*. Current Opinion in Structural Biology, 2003. **13**(6): p. 748-757.
65. Brooks, B.R., et al., *CHARMM - a program for macromolecular energy, minimization, and dynamics calculations*. Journal of Computational Chemistry, 1983. **4**(2): p. 187-217.
66. Faraldo-Gomez, J.D. and B. Roux, *On the importance of a funneled energy landscape for the assembly and regulation of multidomain Src tyrosine kinases*. Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(34): p. 13643-13648.
67. Henzler-Wildman, K.A., et al., *Intrinsic motions along an enzymatic reaction trajectory*. Nature, 2007. **450**(7171): p. 838-U13.
68. Berman, H.M., et al., *The Protein Data Bank*. Nucleic Acids Research, 2000. **28**(1): p. 235-242.
69. Brooks, B. and M. Karplus, *Normal-modes for specific motions of macromolecules - Application to the hinge-bending mode of lysozyme*. Proceedings of the National Academy of Sciences of the United States of America, 1985. **82**(15): p. 4995-4999.
70. Case, D.A., *Normal-mode analysis of protein dynamics* Current Opinion in Structural Biology, 1994. **4**(2): p. 285-290.
71. Hinsen, K., *Analysis of domain motions by approximate normal mode calculations*. Proteins-Structure Function and Genetics, 1998. **33**(3): p. 417-429.
72. Tama, F. and Y.H. Sanejouand, *Conformational change of proteins arising from normal mode calculations*. Protein Engineering, 2001. **14**(1): p. 1-6.

73. Bahar, I., A.R. Atilgan, and B. Erman, *Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential*. Folding & Design, 1997. **2**(3): p. 173-181.
74. Kondrashov, D.A., Q. Cui, and G.N. Phillips, Jr., *Optimization and evaluation of a coarse-grained model of protein motion using X-ray crystal data*. Biophysical Journal, 2006. **91**(8): p. 2760-2767.
75. Yang, L., G. Song, and R.L. Jernigan, *How well can we understand large-scale protein motions using normal modes of elastic network models?* Biophysical Journal, 2007. **93**(3): p. 920-929.
76. Atilgan, A.R., et al., *Anisotropy of fluctuation dynamics of proteins with an elastic network model*. Biophysical Journal, 2001. **80**(1): p. 505-515.
77. Bahar, I., et al., *Vibrational dynamics of folded proteins: Significance of slow and fast motions in relation to function and stability*. Physical Review Letters, 1998. **80**(12): p. 2733-2736.
78. Franklin, J., et al., *MinActionPath: maximum likelihood trajectory for large-scale structural transitions in a coarse-grained locally harmonic energy landscape*. Nucleic Acids Research, 2007. **35**: p. W477-W482.
79. Kim, M.K., R.L. Jernigan, and G.S. Chirikjian, *Efficient generation of feasible pathways for protein conformational transitions*. Biophysical Journal, 2002. **83**(3): p. 1620-1630.
80. Maragakis, P. and M. Karplus, *Large amplitude conformational change in proteins explored with a plastic network model: Adenylate kinase*. Journal of Molecular Biology, 2005. **352**(4): p. 807-822.
81. Zheng, W., B.R. Brooks, and G. Hummer, *Protein conformational transitions explored by mixed elastic network models*. Proteins-Structure Function and Bioinformatics, 2007. **69**(1): p. 43-57.
82. Zheng, W. and M. Tekpinar, *Large-scale evaluation of dynamically important residues in proteins predicted by the perturbation analysis of a coarse-grained elastic model*. BMC Structural Biology, 2009. **9**.
83. Petrone, P. and V.S. Pande, *Can conformational change be described by only a few normal modes?* Biophysical Journal, 2006. **90**(5): p. 1583-1593.
84. Hamilton, D.H., et al., *Large cooperativity in the removal of iron from transferrin at physiological temperature and chloride ion concentration*. Journal of Biological Inorganic Chemistry, 2004. **9**(8): p. 936-944.
85. Blaber, M., et al., *Alanine scanning mutagenesis of the alpha-helix-115-123 of phage-T4 lysozyme - Effects on structure, stability and the binding of solvent* Journal of Molecular Biology, 1995. **246**(2): p. 317-330.
86. Min, W., et al., *Fluctuating enzymes: Lessons from single-molecule studies*. Accounts of Chemical Research, 2005. **38**(12): p. 923-931.
87. Baysal, C. and A.R. Atilgan, *Elucidating the structural mechanisms for biological activity of the chemokine family*. Proteins-Structure Function and Genetics, 2001. **43**(2): p. 150-160.
88. Sacquin-Mora, S. and R. Lavery, *Modeling the Mechanical Response of Proteins to Anisotropic Deformation*. Chemphyschem, 2009. **10**(1): p. 115-118.
89. Zheng, W.J. and B. Brooks, *Identification of dynamical correlations within the myosin motor domain by the normal mode analysis of an elastic network model*. Journal of Molecular Biology, 2005. **346**(3): p. 745-759.
90. Fiorin, G., et al., *Unwinding the helical linker of calcium-loaded calmodulin: A molecular dynamics study*. Proteins-Structure Function and Bioinformatics, 2005. **61**(4): p. 829-839.

91. Wriggers, W., et al., *Structure and dynamics of calmodulin in solution*. Biophysical Journal, 1998. **74**(4): p. 1622-1639.
92. Yang, C., G.S. Jas, and K. Kuczera, *Structure and dynamics of calcium-activated calmodulin in solution*. Journal of Biomolecular Structure & Dynamics, 2001. **19**(2): p. 247-271.
93. Likic, V.A., et al., *A statistical approach to the interpretation of molecular dynamics simulations of calmodulin equilibrium dynamics*. Protein Science, 2005. **14**(12): p. 2955-2963.
94. VanderSpoel, D., et al., *Bending of the calmodulin central helix: A theoretical study*. Protein Science, 1996. **5**(10): p. 2044-2053.
95. Chakrabarti, B. and A.J. Levine, *Nonlinear elasticity of an alpha-helical polypeptide: Monte Carlo studies*. Physical Review E, 2006. **74**(3).
96. Project, E., et al., *A molecular dynamics study of the effect of Ca²⁺ removal on calmodulin structure*. Biophysical Journal, 2006. **90**(11): p. 3842-3850.
97. Yang, L. and Y.Q. Gao, *An approximate method in using molecular mechanics simulations to study slow protein conformational changes*. Journal of Physical Chemistry B, 2007. **111**(11): p. 2969-2975.
98. Alexandrov, N. and I. Shindyalov, *PDP: protein domain parser*. Bioinformatics, 2003. **19**(3): p. 429-430.
99. Humphrey, W., A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics*. Journal of Molecular Graphics & Modelling, 1996. **14**(1): p. 33-38.
100. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD*. Journal of Computational Chemistry, 2005. **26**(16): p. 1781-1802.
101. Darden, T., et al., *New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations*. Structure with Folding & Design, 1999. **7**(3): p. R55-R60.
102. Andersen, H.C., *RATTLE - a velocity version of the SHAKE algorithm for molecular-dynamics calculations*. Journal of Computational Physics, 1983. **52**(1): p. 24-34.
103. Yilmaz, L.S. and A.R. Atilgan, *Identifying the adaptive mechanism in globular proteins: Fluctuations in densely packed regions manipulate flexible parts*. Journal of Chemical Physics, 2000. **113**(10): p. 4454-4464.
104. Baysal, C. and H. Meirovitch, *New conformational search method based on local torsional deformations for cyclic molecules, loops in proteins, and dense polymer systems*. Journal of Chemical Physics, 1996. **105**(17): p. 7868-7871.
105. Baysal, C. and H. Meirovitch, *Efficiency of the local torsional deformations method for identifying the stable structures of cyclic molecules*. Journal of Physical Chemistry A, 1997. **101**(11): p. 2185-2191.
106. Atilgan, C., O.B. Okan, and A.R. Atilgan, *Network-Based Models as Tools Hinting at Nonevident Protein Functionality*, in *Annual Review of Biophysics, Vol 41*, D.C. Rees, Editor. 2012. p. 205-225.
107. Kantardjiev, A.A. and B.P. Atanasov, *PHEPS: web-based pH-dependent protein electrostatics server*. Nucleic Acids Research, 2006. **34**: p. W43-W47.
108. Kantardjiev, A.A. and B.P. Atanasov, *PHEMTO: protein pH-dependent electric moment tools*. Nucleic Acids Research, 2009. **37**: p. W422-W427.
109. Bas, D.C., D.M. Rogers, and J.H. Jensen, *Very fast prediction and rationalization of pK(a) values for protein-ligand complexes*. Proteins-Structure Function and Bioinformatics, 2008. **73**(3): p. 765-783.

110. Gordon, J.C., et al., *H++: a server for estimating pK(a)s and adding missing hydrogens to macromolecules*. Nucleic Acids Research, 2005. **33**: p. W368-W371.
111. Tynan-Connolly, B.M. and J.E. Nielsen, *pKD: re-designing protein pK(a) values*. Nucleic Acids Research, 2006. **34**: p. W48-W51.
112. Farzami, B., A.A. Moosavimovahedi, and G.A. Naderi, *Elucidation of pka values for CA2+ binding-sites in calmodulin by spectrofluorometry*. International Journal of Biological Macromolecules, 1994. **16**(4): p. 181-186.
113. Kesvatera, T., et al., *Focusing of the electrostatic potential at EF-hands of calbindin D-9k: Titration of acidic residues*. Proteins-Structure Function and Genetics, 2001. **45**(2): p. 129-135.
114. Isvoran, A., C.T. Craescu, and E. Alexov, *Electrostatic control of the overall shape of calmodulin: numerical calculations*. European Biophysics Journal with Biophysics Letters, 2007. **36**(3): p. 225-237.
115. Roberts, E., et al., *MultiSeq: unifying sequence and structure data for evolutionary analysis*. BMC Bioinformatics, 2006. **7**.
116. Boschek, C.B., T.C. Squier, and D.J. Bigelow, *Disruption of interdomain interactions via partial calcium occupancy of calmodulin*. Biochemistry, 2007. **46**(15): p. 4580-4588.
117. Blumenthal, D.K. and J.T. Stull, *Effects of pH, ionic-strength, and temperature on activation by calmodulin and catalytic activity of myosin light chain kinase*. Biochemistry, 1982. **21**(10): p. 2386-2391.
118. Dzubiella, J. and J.P. Hansen, *Competition of hydrophobic and coulombic interactions between nanosized solutes*. Journal of Chemical Physics, 2004. **121**(11): p. 5514-5530.
119. Shepherd, C.M. and H.J. Vogel, *A molecular dynamics study of Ca2+-calmodulin: Evidence of interdomain coupling and structural collapse on the nanosecond timescale*. Biophysical Journal, 2004. **87**(2): p. 780-791.
120. Farrell, D.W., K. Speranskiy, and M.F. Thorpe, *Generating Stereochemically-Acceptable Protein Pathways*. Proteins: Struc., Func. and Bioinf., 2010.
121. Farrell, D., et al. *Geometrical Pathways in Biomolecules*. Available from: <http://pathways.asu.edu/>.
122. Smith, D.M.A., T.P. Straatsma, and T.C. Squier, *Retention of Conformational Entropy upon Calmodulin Binding to Target Peptides Is Driven by Transient Salt Bridges*. Biophysical Journal, 2012. **103**(7): p. 1576-1584.
123. Sun, H.Y., D. Yin, and T.C. Squier, *Calcium-dependent structural coupling between opposing globular domains of calmodulin involves the central helix*. Biochemistry, 1999. **38**(38): p. 12266-12279.
124. Strynadka, N.C.J., et al., *Structural details of a calcium-induced molecular switch: X-ray crystallographic analysis of the calcium-saturated N-terminal domain of troponin C at 1.75 angstrom resolution*. Journal of Molecular Biology, 1997. **273**(1): p. 238-255.
125. Mukherjea, P., J.F. Maune, and K. Beckingham, *Interlobe communication in multiple calcium-binding site mutants of Drosophila calmodulin*. Protein Science, 1996. **5**(3): p. 468-477.
126. Ikura, M., et al., *Secondary structure and side-chain H-1 and C-13 resonance assignments of calmodulin in solution by heteronuclear multidimensional NMR-spectroscopy*. Biochemistry, 1991. **30**(38): p. 9216-9228.
127. Mackall, J. and C.B. Klee, *Calcium-induced sensitization of the central helix of calmodulin to proteolysis*. Biochemistry, 1991. **30**(29): p. 7242-7247.

128. Huang, S.L., G.M. Carlson, and W.Y. Cheung, *Calmodulin-dependent enzymes undergo a proton-induced conformational change that is associated with their interactions with calmodulin*. Journal of Biological Chemistry, 1994. **269**(10): p. 7631-7638.
129. Huang, S.L. and W.Y. Cheung, *H⁺ is involved in the activation of calcineurin by calmodulin*. Journal of Biological Chemistry, 1994. **269**(35): p. 22067-22074.
130. Chen, B., et al., *Helix A stabilization precedes amino-terminal lobe activation upon calcium binding to calmodulin*. Biochemistry, 2008. **47**(35): p. 9220-9226.
131. Anthis, N.J., M. Doucleff, and G.M. Clore, *Transient, Sparsely Populated Compact States of Apo and Calcium-Loaded Calmodulin Probed by Paramagnetic Relaxation Enhancement: Interplay of Conformational Selection and Induced Fit*. Journal of the American Chemical Society, 2011. **133**(46): p. 18966-18974.
132. Israelachvili, J., *Intermolecular & Surface Forces*. Second ed. 1992: Academic Press.

List of Publications

S. Negi, **A.O. Aykut**, A.R. Atilgan, C. Atilgan, "Calmodulin Readily Switches Conformation upon Protonating High pKa Acidic Residues," J. Phys. Chem. B, 116, 7145-7153 (2012).

A.R. Atilgan, **A.O. Aykut**, C. Atilgan, "Subtle pH differences trigger single residue motions for moderating conformations of calmodulin," J. Chem. Phys., 135, 155102 (2011).

Atilgan C., **Aykut A.O.**, Atilgan A. R., How a Vicinal Layer of Solvent Modulates the Dynamics of Proteins. Biophysical Journal, 94:79-89 (2010).

Atilgan A.R., **Aykut A. O**, Atilgan C., Non-Equilibrium Fluctuation Theorems, Redundant Paths in Proteins, and Elucidating Conformational Changes by Single-Residue Perturbations. Biophysical Journal, 104: 70A, (2013).

Aykut A. O, Atilgan C. and Atilgan A.R., Molecular Recognition Mechanism of Calmodulin Examined by Perturbation Response Scanning. MRS Proceedings, 1301:mrsf10-1301-oo09-09 (2011).